AD-A240 006

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public repor... hour per response, including the time for reviewing instructions, searching existing data sources, gathering a... ction of information. Send comments regarding this burden estimate or any other aspect of this collection of ... gton Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highw... ment and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE 7 Aug 91

3. REPORT TYPE AND DATES COVERED FINAL 01 Sep 87 to 30 Jun 91

4. TITLE AND SUBTITLE

ON CATEGORIZING SOUNDS

5. FUNDING NUMBERS

GR- AFOSR-87-0353
PE- 61102F
PR- 2313
TA- A6

6. AUTHOR(S)

Dr Gregory R. Lockhead

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Department of Psychology
Duke University
Durham NC 27706

8. PERFORMING ORGANIZATION REPORT NUMBER

AFOSR-TR- 91 0731

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Dr John F. Tangney
AFOSR/NL
Building 410
Bolling AFB DC 20332-6448

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

DTIC
ELECTE
SEP 05 1991
S D
D

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Context is important when people judge sounds, or attributes of sounds, or other stimuli. It is shown how judgments depend on what sounds recently occurred (sequence effects), on how those sounds differ from one another (range effects), on the distribution of those differences (set effects), on what subjects are told about the situation (task effects), and on what subjects are told about their performance (feedback effects). Each of these factors determines the overall mean and variability of response times and response choices, which are the standard measures, when people judge attribute amounts. Trial-by-trial analysis of the data show these factors also determine performance on individual trials. Moreover, these momentary data cannot be predicted from the overall data. The opposite is not true; the averaged data can be predicted from the mementary details. These results are consistent with a model having two simple assumptions: Successive sounds (not just their attributes) assimilate toward one another in memory, and judgments are based on comparisons of these remembered events. It is suggested that relations between attributes, rather than the magnitudes of the attributes themselves, are the basis for judgment.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
(U)	(U)	(U)	UNLIMITED

NSN 7540-01-280-5500

91 9 5 007

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

CATEGORIZING SOUNDS

Gregory R. Lockhead

Department of Psychology

Duke University

Durham, North Carolina 27706

DGREG@DUKEMVS.BITNET

Grant No. AFOSR-87-0353

Final Technical Report

Period covered: 1 October 1987 to 30 June 1991

August 15, 1991

Grant Monitor: John F. Tangney, Program Manager,

Life Sciences Directorate

Prepared for: Directorate of Life Sciences

Air Force Office of Scientific Research

Bolling Air Force Base

Washington, D. C. 20332-6448

ON CATEGORIZING SOUNDS

Abstract

Context is important when people judge sounds, or attributes of sounds, or other stimuli. It is shown how judgments depend on what sounds recently occurred (sequence effects), on how those sounds differ from one another (range effects), on the distribution of those differences (set effects), on what subjects are told about the situation (task effects), and on what subjects are told about their performance (feedback effects). Each of these factors determines the overall mean and variability of response times and response choices, which are the standard measures, when people judge attribute amounts. Trial-by-trial analyses of the data show these factors also determine performance on individual trials. Moreover, these momentary data cannot be predicted from the overall data. The opposite is not true; the averaged data can be predicted from the momentary details. These results are consistent with a model having two simple assumptions: Successive sounds (not just their attributes) assimilate toward one another in memory, and judgments are based on comparisons of these remembered events. This holds for judgments of multidimensional stimuli as well as for judgments of unidimensional stimuli. When two dimensions varied between trials but only one was judged, variations of the nominally irrelevant dimension interfered with judgments of the relevant dimension. Furthermore, the magnitude of this effect was greater when the irrelevant dimension varied by larger amounts. For example, pitch judgments depend on whether and by how much loudness changes between trials. Combined with the literature, the results allow the suggestion that continuing to search for an underlying psychophysical scale may not be productive. A different approach is suggested. The traditional approach uses methods adopted from classical physics to examine how people process attributes of objects. The suggested alternative is based, instead, on biological and psychological considerations of how people process objects in various environments. This view is based on the fact that sounds, like most other stimuli used in psychophysical studies, are integral. This means the entire stimulus, rather than its attributes, is processed initially. According to the model, successive stimuli are compared in memory and subjects then judge their attributes. The loudness and pitch judgments reported here are consistent with the interpretation. These and many other data are not consistent with assumptions made by classic scaling models. It is concluded that Fechner's Law, Stevens' Law, and all related psychophysical scaling models are wrong or incomplete. Accordingly, they are rejected. It is suggested that relations between attributes, rather than the magnitudes of the attributes themselves, are the basis for judgment. To present the argument in this technical report, a paper submitted for publication is reproduced here.

G. R. Lockhead

ON CATEGORIZING SOUNDS

The general goal of the AFOSR project was to better understand how objects are identified. The specific goals supported by AFOSR were to evaluate a model of sequence effects in univariate tasks and to learn if the model generalizes to situations in which tones vary in more than one way from trial to trial, i.e., to multidimensional stimuli. Much of the resulting research is summarized in the following paper which has been submitted for publication. The paper is titled: **PSYCHOPHYSICAL SCALING: Judgments of attributes or objects?**

Abstract

Psychophysical scaling models of the form R = f(I), where R is the response and I is some intensity of an attribute, all assume people judge amounts of an attribute. With simple biases excepted, most also assume judgments are independent of space, time, and other features of the situation than the one being judged. Many data support these ideas: Magnitude estimations of brightness (R) increase with luminance (I). Nevertheless, I conclude the general model is wrong. A reason from the stabilized retinal image literature is that nothing is seen if light does not change over time. A reason from the classification literature is that dimensions often *combine* to produce emergent properties that cannot be described by the elements in the stimulus. Other reasons are discussed. These various effects cannot be adjusted for by simply expanding the general model to the form $R = f(X_1, X_2, X_3, \ldots, X_n)$ because some factors do not combine linearly. The proposed alternative is that people initially judge the entire stimulus, the object in terms of its environment. This agrees with the constancy literature which shows that objects and their attributes are identified in terms of their relations to other aspects of the scene. This fact, that the environment determines judgments, is masked in scaling studies where the standard procedure is to hold context constant. Consider a typical brightness study where different lights are presented on the same background on different trials. The essential stimulus for the observer might be the intensity of the light, or it might be a difference between the light and the background. The two are perfectly confounded. This issue is examined for audition. It is shown that judgments of the loudness of a tone depend on the amount by which that tone differs from the previous tone in both pitch and loudness. To judge loudness (and other attributes) it is suggested that people first process the stimulus object (the whole or integral thing) in terms of differences between it and other aspects in the situation, and only then assess the feature of interest. The summary conclusion is psychophysical judgments will be better interpreted by theories of attention based in biology or psychology, than by theories that follow Fechner's lead and are based in classical physics.

PSYCHOPHYSICAL SCALING: Judgments of attributes or objects?

Psychology has long searched for a general psycho-physical law to relate the system output to its input. For example, might brightness double when the intensity of the light doubles? Fechner proposed the first such function that was widely accepted. He said equal stimulus ratios correspond to equal sensation differences. This is Fechner's Law.

A hundred years later, Stevens (1961) published "To honor Fechner and repeal his law" with the goal of replacing Fechner's law with his Law. He said "equal stimulus ratios produce equal subjective ratios" (1957, p. 153). This is Stevens' Law.

Stated formally, these Laws are:

$$R = k \log(I/I_0), \quad \text{Fechner' Law, [1], and}$$

$$R = aI^b, \quad \text{Stevens' Law, [2],}$$

where R is subjective magnitude, I is the physical magnitude or intensity of the scaled attribute, I_0 is absolute threshold, and $\underline{a}$, $\underline{b}$, and $\underline{k}$ are constants.

Stevens' attempt to replace Fechner was influential but not completely successful. This has been a lively issue. Many papers have examined the laws, compared them, and suggested alternatives (summaries in Bolanowski & Gescheider, 1991; Falmagne, 1985; Laming, 1988; Luce & Krumhansl, 1988; Poulton, 1989; others). Still, there is no general agreement that one or the other is correct. In an attempt to reach a resolution, Lester Krueger suggested both are partially correct and the solution to "the true psychophysical scale ... lies halfway between that of Fechner ... and that of Stevens" (1989, p. 251). Perhaps so, although compromise is not commonly long lived in science. I suggest the reason there is no consensus is that neither model describes what is involved when people judge magnitudes.

Equations [1] and [2] describe data collected under particular conditions. Causes of the outcomes are not known. <u>Why</u> are more intense lights judged to be brighter? The implicit answer, that they have more energy or intensity or luminosity or usable power is incomplete or wrong. Several reasons it is wrong are presented ahead. One comes from the stabilized-retinal-image literature; a light is not seen if it does not change over time at the eye, no matter how intense it is (Arend & Timberlake, 1986; references there). Thus time is involved. Other reasons include 1) Relations between attributes, not the amount of an attribute, are the essential stimulus for the subject, 2) what occurred prior to the stimulus determines its perception and response, and 3) attributes often combine such that new features emerge and these cannot be deduced from the attributes. Because of such facts, Equations [1] and [2] are not sufficient to account for the findings, and a linear expansion from their form,

$$R = f(I) \qquad [3]$$, to the more general form

$$R = f(I, X_1, X_2, X_3, \ldots X_n) \qquad [4],$$

where X_i are factors that affect the response in addition to I, is also not sufficient. Some other approach is needed. The aims of this article are to repeal all such psychophysical scaling laws and to suggest an alternative view of psychophysical judgments.

My fundamental argument is that psychophysical scaling models assume stimulus attributes are judged independently of their environment, and this is wrong. The stimuli used in most psychophysical tasks are <u>integral</u>. People cannot process one attribute of an integral stimulus independent of other attributes (Lockhead, 1966, 1970, 1972, in press; Garner, 1974; Shepard, 1991).

These remarks are not embarked upon without trepidation. Psychophysical scaling has a rich history and has even become an industry that serves many needs. Too, psychophysical scales are certainly correct at some level. As anyone who has conducted a magnitude estimation experiment knows, the data that are produced cluster closely around the power function according to Stevens' Law; averaged responses are commonly linear with the physical intensities of the attributes, when both measures are plotted on logarithmic scales. Furthermore, this result is not unique to the particular method. Various scales (category, magnitude, neuroelectric, summated jnd) are often linearly related when appropriate adjustments are made.

Essentially all current psychophysical scaling models take the form of Equation [3] or [4]. This includes functional measurement theory that was constructed by Anderson (1981) to avoid some scaling difficulties and conforms to Equation 4. This also includes the relation theory of Krantz (1972) and Shepard (1981). Their basic axiom is that numbers representing the physical intensities of stimulus attributes are mapped onto sensory continua and are then related (Falmagne, 1985, p. 311). This assumes the subject's stimulus is attribute intensity, just as in other scaling models, and further assumes that encoded intensity is related to the memory of the previously encoded intensity.

All psychophysical scaling models assume:

<u>Assumption I</u>. The subjective magnitude of an attribute of a stimulus is some function of the physical magnitude of that attribute.

They also assume, or allow the reader to imply, that effects of context can be controlled or can be removed from data such that:

Assumption II. Attribute judgments are independent of other attributes,

Assumption III. Attribute judgments are independent of spatial context, and

Assumption IV. Attribute judgments are independent of time.

Stevens made Assumptions I, II, III, and IV explicit. While recognizing discrepancies and increased variability in data when people matched the brightnesses of differently colored lights, as compared to matching the brightnesses of lights with the same hue and saturation, he stressed that people have "the ability to separate out of a complex configuration one single aspect and to compare that aspect with the same aspect abstracted from another configuration" (1975, p. 66). Consistent with this, Krueger (1989, p. 264) concluded that when various scaling methods are used for the same stimulus dimension, experimental adjustments to the resulting data "were successful in removing sources of systematic error or bias ... and that the common function provides a true point-by-point mapping of physical magnitude, I, into subjective magnitude".

It should immediately be noted that Assumptions I-IV only concern psychophysical scaling. They do not apply to all research in psychophysics. Some researchers who use R = f(I) do so as a shorthand to report data but do not do seek an underlying psychophysical scale. Prominent examples include Marks (1974) who has argued that psychophysical judgments are multidimensionally determined, and Laming (1985, 1988) who has concluded that "experiments which have formerly been claimed to measure internal sensations can be adequately understood by reference to the physical level of description alone, without any suppositions about internal machinery" (1988, preface). Indeed, Falmagne (1985; Gescheider, 1988, made a related point) decided there are two classes of researchers who report psychophysical scales. One seeks to uncover the psychophysical law. For them, biases and other context effects in data are to be removed in order to reveal the correct scale and the above assumptions are relevant. The other adopts some scale only as a convenience when reporting data and as an aid for understanding sensory processing. Falmagne's view, with which I agree, is that it is difficult to support belief in a particular scale and "there is no strong argument that progress in sensory research calls for standard scales" (1985, p. 322).

Nonetheless, many people continue to search for scales. One reason may be a continuing belief in 19th century foundations of experimental psychology. Many people then thought there is a psychological world that is independent of the physical world and, since we nonetheless function in relation to the external world, some way to map one world onto the other was needed (Boring, 1950). Psychophysical scaling suggested an answer.

The search for a true psychophysical scale has carried with it the, often unstated, Assumption II (above) that stimulus attributes are abstracted and judged separately from the rest of the object or its environment. For example, when people are asked to judge the brightness of a disc, it is assumed they do so. I argue they do not, at least not in any direct way. Instead, I suggest people perceive the disc as part of the environment, and then assess its brightness in terms of the situation.

This substitute view is based in part on a common argument concerning natural selection: It is more important for organisms to identify objects than to measure the intensities of their attributes, and perception evolved accordingly. Attributes can vary independently of objects and so do not reliably predict objects in the ordinary world: The amount of light coming from the fur of a tiger in shadow is less than that coming from the fur of the same tiger in sunlight. While we may note the darkness or lightness of the situation, it is more important to know there is a tiger. I propose that is what perception accomplishes.

Intensity versus intensity differences in time and space

I show here that the magnitude of a physical attribute is not the appropriate dependent variable for a general scaling model. This does not mean no such model is possible, but it does call into question any scaling model based on attribute intensity. Using the brightness of a stimulus disc as an example, this section summarizes some previous work showing that the intensity of the disc is not generally available to subjects for judgment.

To introduce the argument, it is useful to note how assumptions about psychophysical scaling reflect the approach that helped establish classical physics. This comparison, which I have often used in lectures but which is now borrowed from Allik who published a similar one (1989, 267-268), is as follows: The volume of a gas increases linearly with its temperature (when pressure is held constant and degrees Kelvin is measured), and the apparent amount of gas also increases with temperature. Thus, the same general function holds in an algebraic expression by physicists and in phenomenology; volume (a physical measure) and apparent amount (a psychological measure) both rise with temperature. While this is now trivial, surely when people were founding classical physics it was useful to have appearances that agreed with equations relating physical attributes.

Early psychology also capitalized on this correlation between physical amounts and appearances. To increase the intensity of a tone means to increase the amplitude of the soundwave <u>and</u> to increase its loudness. Just as for the volume, temperature, and apparent amount of a gas, the acoustic energy, amplitude, and loudness of a tone are also correlated algebraically and phenomenally. Many such parallels between phenomenology and algebra facilitated the development of classical physics by providing the field with face validity. Those parallels also gave credence in psychology (or philosophy) to a <u>psycho</u>-physical model that had the same form as <u>physical</u>-physical models in physics. In both cases, physical measures were consonant with appearances when physical and phenomenal properties covaried.

Such equations often work well, at least within limits. However, this does not necessarily mean underlying or causative factors are captured by those equations. Indeed, the classic physical model to describe the action of gases is wrong in general and has been replaced with thermodynamics. A theory initially based on directly observable properties was replaced by a theory based on underlying properties. Eventually, theories of psychological scaling based on observable properties might also be replaced by theories based on underlying properties. I think that one candidate for replacement is the class of models that relate the physical intensity of an attribute to the phenomenal magnitude of that attribute.

<u>Intensity or intensity change?</u> Although the particular arguments in this section are restricted to brightness, more general statements can be made and are intended. It is argued that no static model relating intensity to brightness is sufficient. This is because, in order to describe the facts, it is necessary to write equations in terms of changes-over-time in the intensity of an attribute, and time is not involved in any of the models. This distinction changes the implied sensory or perceptual process from one in which intensity (a physical measure) causes brightness (a psychological measure) to one in which intensity is no more than indirectly related to brightness and for which the implied physiological process is different from what would be speculated on the basis of intensity alone.

Static models of brightness imply that the number of photons per unit time (quantitative differences associated with wavelength and with receptor sensitivity are not essential to arguments here) determine brightness. This cannot be correct. Nothing is seen if intensity at the eye does not change over time. Perhaps the most dramatic demonstration of this is that stabilized retinal images disappear (Krauskopf, 1963). <u>Without changes in the light over time at the receptors we are blind.</u>

How is it then that data from psychophysical tasks beautifully support the classic psychophysical functions? How is it that brightness (the psychological measure) increases with intensity in those studies? The answer, I suggest, is that amount of intensity, <u>I</u>, has regularly been confounded in scaling tasks with changes-in-amount-of-intensity, <u>delta I</u>, and delta I is the causal variable.

Consider how psychophysical studies of brightness are commonly conducted. Usually after some adaptation period, a luminous disc is presented out of darkness or out of a uniform field of intensity. This stimulus disc has different intensities on different trials. The easily replicated finding is response numbers are larger for stimuli having more intensity (more energy in that region per unit time). This is true. The common inference is that the brightness level is caused by the amount of intensity. This is not true.

Consider what happens when the stimulus is first turned on. There is then a change at the eye. The intensity level shifts from that of the background to that of the stimulus. On trials that the intensity level is large, the amount of this change in intensity over time is also large. Thus, it cannot be known from such studies whether responses are due to amount-of-intensity or to amount-of-change-of-intensity. The two are confounded. Three aspects of this confound are discussed next.

<u>Brightness of a flashed light.</u> One procedure to study brightness as a function of stimulus duration is to flash a light of fixed intensity (fixed energy per unit area) for different durations. Its brightness is matched by having the subject manipulate the intensity of another light that is maintained on until the subject is satisfied that the two lights have the same appearance, except for their durations. The luminance of the matching light is recorded as the brightness of the flashed light.

If brightness were due only to intensity, then this matching luminance should be independent of flash duration. It is not. The upper panel of Figure 1 shows that brightness (matching luminance) increases with duration up to some value, then decreases as the flash duration increases further, and seems to level off as the flash duration increases even further (Arend, 1970) This is consistent with earlier data (Aiba & Stevens, 1964) and with the theoretical description adapted from Anglin and Mansfield (1968) and shown in the lower panel of Figure 1. Brightness increases linearly with flash duration up to some critical duration, t_c, decreases as the flash duration increases further, and is independent of time at greater durations. This pattern holds qualitatively across different intensities of the flashed disc. The quantitative differences are that when the intensity of the flashed light is made greater, then brightness is greater at all flash durations and t_c is briefer.

---Figure 1---

When the dependent variable for such studies is detection of light rather than matching luminance, the data demonstrate Bloch's law (1885). Lights that are equal in total intensity are equally detectable up to some critical duration. $IT = k$, for $T < t_c$, where *I* is stimulus intensity, T is stimulus duration, k is absolute threshold, and t_c is generally less than 250 msec.

A common interpretation of such findings is that, up to t_c, the total amount of intensity within some time window determines brightness. Since the stimulus light is flashed out of darkness or out of a fixed background in such studies, it is equally possible, instead, that the total amount of change-in-intensity within some integrating period determines brightness. This is the guess here.

Figure 1 also shows the Broca-Sulzer effect (1902; Katz 1964); brightness decreases when stimulus duration increases beyond t_c. This decrease in brightness with increased flash duration is sometimes interpreted to indicate the onset of inhibition. This interpretation is not needed for the current thesis and I do not comment on it.

Since matching luminance has many different values over time for the same intensity light (same energy per unit area), duration is involved in brightness. Concerning Stevens' Law, this means the exponent of the power function must be different when the same intensities are presented for different durations. J. C. Stevens & Hall (1966) showed this is indeed the case. Anglin and Mansfield (1968) then showed that Stevens' model can describe such data but "a different exponent is required at short durations from that required at long durations" (p. 161). Brightness is not a function of only intensity, at least not for short durations.

<u>Brightness of a steady light.</u> What about long durations? According to Figure 1, brightness is then independent of duration. Are not standard psychophysical models then appropriate? The answer is again no. At long durations there is confounding due to two other facts: The stimulus has contours or edges, and the eye and body are in constant motion. Because of these movements, the stimulus contours are regularly moved onto and then off of some receptors. This means light is regularly moved onto and then off of receptors that are near the stimulus edges, even though the light itself is steadily on.

One way to decontaminate possible effects of intensity from effects of intensity-changes caused by the movement of contours is to eliminate all edges in the stimulus. This can be done by flooding the eye uniformly with light. This produces a ganzfeld or uniform field. Now there are no contours to be moved across the retina when the eye moves. Thus, there are no temporal changes in luminance at the eye (as long as the subject does not blink or otherwise interfere with the situation) when the light is maintained on. The result is the light is seen when it first floods the eye and then fades and disappears (Kelly, 1979, and several earlier demonstrations).

This might lead to the speculation that it is luminous contours, not just temporal changes in intensity at the receptors as proposed here, that are needed for vision to endure beyond the initial stimulation. Testing this hypothesis leads to asking if it is possible to deconfound intensity from intensity-change when there are spatial contours at the eye. This answer would immediately be yes if the eye could harmlessly be prevented from moving. Unfortunately, this is difficult to accomplish. Attempts by John Monahan and me to anesthetize extraocular muscles in order to stop eyemovements resulted in also anesthetizing the optic nerve. Then no vision tests could be conducted.

Fortunately, as already noted, a more successful method of stopping movements of a contoured image is available. Rather than stopping the eye, this solution is to stabilize the image on the retina by, using mechanics and optics, moving the stimulus as the eye moves. In this situation a luminous pattern can be kept on the same receptors over time even though the eye moves about in the orbit and even though the head and body move.

When such a stabilized-retinal-image technique is used, brightness is not maintained at long durations. Rather, the stimulus pattern is seen when it is first turned on (an intensity change over time) and then the visual world disappears (Yarbus, 1967). Yarbus guessed the disappearance occurs within 2 sec. That estimate includes the time required for the observer to note the disappearance and to then report it verbally.

Introspection suggests the disappearance occurs even more quickly. To see this, hold a penlight at the side of your eye and wiggle the light. With practice, you will see the entopic shadows of your retinal blood vessels that lie between the light and your photoreceptors (Campbell & Robson, 1961; Sharpe, 1972). Those shadows are stabilized on the receptors under normal viewing because they move as the eye moves. The wiggling light moves them across receptors. This produces intensity changes over time at the retinal surface and the shadows are seen.

When you stop wiggling the light, the shadows disappear. It seems to me that the disappearance occurs in less than 1/2 second. Perhaps it happens between 100 and 250 msec, which is about when the Broca-Sulzer decay begins. Whether or not this is so, the image is seen when the light is moved and it disappears when the light is no longer moved, even though the shadows are still on the retina. Such findings mean that neither light (luminous intensity) nor light differences (a luminous pattern) at the eye are sufficient for the image to be reported. Indeed, they are apparently not even sufficient for brightness to be reported (Arend & Timberlake, 1986). Although there are technical difficulties in producing a perfectly stabilized image and thus in clearly proving the conclusion (e.g, pulsations in the retina due to the heart beat cause images changes), the best current estimate is that brightness is zero when an illuminated stimulus, with or without spatial contours, does not change over time on the retina.

This does not mean the visual world is then black. Black is only reported when there is a seen brightness to provide contrast. When there is no contrast, the appearance is a uniform dark gray instead of black. Hering (1964) called this the eigengrau or the gray of the eye, where, again, intensity and appearances are not related.

The most common reaction by students in my classes, when they are told that stabilized images disappear, is to say the eye adapted to the light. This interpretation can be no more than comforting. It does not explain why the page you are reading does not disappear. The reason it does not vanish is that eye and body movements assure the image is not fixed on the retina for longer than a fraction of a second.

It has been instructive to then remind students about the experimental procedure used to produce the data in Figure 1. What about the steadily illuminated disc, the standard, that matches the appearance of the test disc flashed for a long duration? Why does that not disappear? The stabilized retinal image data indicate the reason is not that the standard is fixed in time. Rather, it is because eyemovements produce temporal changes in the standard at the eye. Otherwise, it would vanish.

Certainly, some visual adaptation does occur over time and the brightness level of a stimulus does depend on the adapted state of the visual system. However, the reason stabilized images disappear is not adaptation or fatigue. The disappearance is too fast for that. Instead, light is not the stimulus for vision. Maintained-on fibers and related physiological facts notwithstanding for these psychophysical measures, while light is necessary for vision the critical parameter is change in light intensity over time (see similar suggestions in Arend, Buehler, & Lockhead, 1971; Arend & Timberlake, 1986; Krauskopf, 1963; Laming, 1985, 1988; Lockhead, 1988; Yarbus, 1967; others). Rather than being due to the steady presentation of intensity, <u>brightness and thus vision are a result of a continuing sequence of snapshots of intensity</u>.

<u>Brightness and remote contours.</u> The above conclusions are that intensity and intensity differences are not sufficient, and intensity change over time is necessary for there to be vision. Brightness levels were not discussed. This section shows that the amount of intensity or intensity change at any region of the eye is not sufficient to predict the brightness level at that region.

Simultaneous brightness contrast is perhaps the best known demonstration of the fact that the amount of intensity at a location is not sufficient to predict the brightness there. Simultaneous-brightness-contrast describes the fact that a fixed intensity disc has a different brightness when it is seen against a different background. The same patch of light appears brighter when the surround on which it is viewed is made less intense (Heinemann, 1950). An intimatelv related fact is that the difference limen (how much an intensity must be changed for a light to be seen as different) also depends on the background on which the lights are viewed (Brysbaert & d'Ydewalle, 1989; Graham & Kemp, 1938). Any complete psychophysical scaling model must account for such facts. None does beyond treating the intensity of the background as a parameter.

To show why an intensity difference is not sufficient to determine brightness at that location, consider stimuli where the intensity across space changes gradually, rather than abruptly as with more commonly used step-functions. An example is the Craik-O'Brien-Cornsweet effect. This shows that physically identical luminous areas, each containing small luminous gradients, are different in brightness when they are separated by luminous steps. An extension of this effect is shown in Figure 2 (cf. Arend, Buehler, & Lockhead, 1971). The top portion of the Figure 2 shows a black and white (construction papers) distribution that was pasted on a disc and spun rapidly. The result is the photograph in the bottom portion of the figure. The stimulus has three identical Craik-O'Brien-Cornsweet gradients, with a step function (a bar) superposed on the outer and the inner gradients.

The black/white ratios for the three Craik-O'Brien-Cornsweet distributions are identical to one another. When the disc is spun, the black/white ratios are the same in the three regions. Nonetheless, these areas differ in brightness. The outer one is darkest, the middle one is intermediate, and the inner one is lightest. Furthermore, the two bars, which are identical in luminance and are superposed on identical surrounds, also differ in brightness. Hence, the brightness of the bars is not determined by their intensity or by their intensity differences with the surround. It is determined by the entire layout. Since the entire spatial configuration must be taken into consideration to describe brightness, Assumption III, that attribute judgments are independent of spatial context, is rejected.

---Figure 2---

Conclusion.

Intensity does not determine brightness. Assumption I is wrong and R = f(I) is rejected as a general psychophysical scaling equation. This means that Fechner's Law, Stevens' Law, and all other such models must be repealed or modified.

Only brightness was discussed in reaching this conclusion. One might ask "but what about roughness and hardness and loudness and other intensive dimensions, and what about nonintensive dimensions, such as pitch and hue and orientation?" That is, might brightness be special? I cannot prove that <u>no</u> physical attribute is directly judged. But that is not an issue. One exception is sufficient to disprove a general rule. Brightness does that.

Moreover, brightness is not the only attribute subject to the criticism that the stimulus is not simply intensity. Another is loudness. Loudness results from compression and rarefaction of air molecules over time. Steady pressure is not heard and so intensity (force per unit area) is again not the stimulus. Change in pressure over time is needed. This is so well known that it may seem silly to even mention that temporal change, frequency, is an integral aspect of sound. I note this here only to further indicate the importance of intensity-changes-over-time.

When these pressure changes over time are sufficient to produce loudness, pitch and timbre then also occur. All these attributes are required for one of them to exist. The parallel is also true in vision. When changes in luminous over time are sufficient to produce brightness, saturation and hue are also occur. These perceptual dimensions are integral with one another. Similar observations are available for other senses.

This rejection of intensity-based scaling models does not prove that no psychophysical scale exists. That requires proving the null hypothesis, which is logically impossible. But if there is a psychophysical scale, then it must be based on the first or second derivative, or some other function, of energy with respect to time and space. Discovery of such a scale would have important but different implications for psychology, physiology, engineering, and theory than do classic models. This is because it would implicate different mechanisms than are suggested by classic models.

This argument concerning the null hypothesis can and should be turned around. Rather than waiting to be proven wrong, it is the task of the theorist who proposes a psychophysical scale, or who proposes anything else, to prove its existence. Extensive efforts notwithstanding, this has not been accomplished for psychophysical scaling. The following section on context demonstrates this will be difficult to accomplish even if some scaling law is true.

Context.

The classical scaling laws were obtained by varying one attribute while holding other factors constant. People judged, for example, the loudnesses of tones having different intensities but the same frequency, duration, and apparent location. Such experimental control is a key ingredient of the scientific method. However, once some invariance is uncovered, e.g., loudness and intensity are linearly related along some scales, it becomes important to examine factors that might be confounded with those measures in order to better understand the cause of the invariance.

In this section, I examine effects of factors other than the attribute to be judged. To anticipate, I will conclude that so many things affect judgments that any prospect of removing their effects to reveal a true, underlying function is remote. Psychophysical scaling requires much more from the subject than the detection or discrimination of an attribute followed by a simple assignment of numbers or some other match to that attribute. One of the first demonstrations of such context effects is a 1954 study of half-loudness judgments in which Wendell Garner concluded that observers "do not seem able to describe sensory magnitudes with a scale of numbers" (1954, p.

224). Rather, responses "seem to be more influenced by the context of stimuli provided him [the subject] than they are by any loudness scale in his sensorium" (Garner, 1958, p. 1007 .

The sizes of these effects of context are also remarkable; response variance in scaling tasks is typically 100 times as great as it is in threshold discrimination tasks (Laming, 1988). It is not only that scaling data are highly variable. There are also large performance differences associated with the particular task and with the particular stimulus set studied (cf. Poulton, 1989). These are also not trivial effects that can safely be averaged over or ignored for scaling purposes. Examples reported from my laboratory alone include the following: Changes in the stimulus range have affected judgments by a factor of six (also see Teghtsoonian, 1973), differences in stimulus sequence and differences in the experimental task have both affected responses by 75% of the response range, and giving or not giving feedback and manipulating how stimulus attributes are combined have both shifted accuracy from near chance to near perfect (Lockhead, 1970, 1984, in press; Lockhead & King, 1988). Furthermore, such context effects do not simply add a constant to judgments. For example, stimulus range and scaling procedures "influence not only overt responses scales, but measures of underlying intensity processing" as well (Algom & Marks, 1990, abstract).

Such situational effects have theoretical consequences. It would be foolish to try to evaluate a psychophysical scaling model without at least controlling or measuring these effects. For example, the slope of the power function in magnitude estimation data varied by a factor of three when the response range was manipulated (King & Lockhead, 1981), and there is no basis for knowing which range is the "correct" range to use in producing a scale. Since there are many context effects, some of which interact, it may not be possible to remove them (but see Anderson, 1981; Birnbaum, 1982; DeCarlo & Cross, 1990) in order to reveal an underlying true psychophysical scale of attribute intensity. The following studies further reveal the difficulty of measuring an underlying psychophysical scale.

I. Judgments of univariate stimuli.

Univariate stimuli differ from one another along only one physical dimension, such as wavelength, extent, weight, or intensity. Some context effects in judgments of univariate stimuli are summarized in this section. Although these findings have been reported previously, some detail is given for the reader who has not encountered this literature.

Sequence effects. The study that first directed my attention to context effects in psychophysical judgments was one in which people made absolute judgments of the intensities of ten tones that varied only in amplitude (Holland & Lockhead, 1968). We asked people to identify each randomly presented tone with a numeral (1-10). The subjects were given feedback (1-10) after each response.

Figure 3 shows the mean response error as a function of the stimulus that occurred k trials earlier. For example, the top point in the figure shows that, compared to the overall average, the average response was about 0.4 category units larger when stimulus #9 or #10 occurred on the prior trial (k = 1). That is, stimuli were overestimated when the prior stimulus was large. Similarly, stimuli were underestimated when the prior stimulus was small, a #1 or #2. In general, judgments tended to be similar to value of the prior trial. This is known as assimilation. Figure 3 additionally shows that judgments tended to be different from stimuli that occurred earlier in the sequence (k = 2 to 5 or more). This is known as contrast. Assimilation and contrast had each been observed previously in data (Helson, 1964), but this is the first time both effects were seen in the same data set.

Since judgments assimilate toward the prior trial and contrast from earlier trials, judgments are not independent of time or events occurring over time. Furthermore, the magnitude of this assimilation depends on the inter-trial interval (Holland, 1968). Since judgments are not independent of time, Assumption IV is rejected.

---Figure 3 here---

Assimilation occurs in responses **and** in memories of stimuli. The contrast seen in Figure 3 is largely associated with response adjustments that are made by the subjects to correct for errors that had been caused by assimilation (King, 1980; Staddon, King, & Lockhead, 1977) and contrast is not considered further in this paper. Assimilation is considered further to help explain what is involved when people identify stimuli and their attributes.

Perfectly locating the source of assimilation or contrast or any other psychophysically measured effect, whether in physiology or in a psychological process model, may not be possible. However, some determinations can be made. One might ask if assimilation is due to sensory adaptation, sensory fatigue, short term memory, or response bias. If there is only one source and it is one of these, this answer is response bias. This is because there is assimilation in guessing studies in which there are no stimuli (Ward & Lockhead, 1971). Sensory effects and stimulus memories could not be involved in those data because there were no differential sensory stimulations and there were no stimuli to be remembered.

The fact that assimilation occurs in response systems does not rule out the existence of assimilation in perception or in the memory of stimuli as well. Assimilation may have many sources. To examine this, we asked people to judge the relative intensities of successive tones (Lockhead & King, 1983) in a successive-ratios-judgment task. The stimuli were 30 auditory sine waves spaced in 1 dB steps and presented in random order for many trials.

As an example of the results, which were quite general, consider when the 74 dB tone was presented on successive trials. The ratio between those tones is $S_N/S_{N-1} = 74/74 = 1$. Thus, the response in this successive-ratios-judgment task should be "1". But "1" was rarely given. Instead, averaged responses were greater than 1 if the tone just before these two 74 dB tones (S_{N-2}) was less than 74 dB, and less than 1 if S_{N-2} was greater than 74 dB. Our interpretation is this occurs because the first 74 dB tone (S_{N-1}) assimilated in memory toward the tone just before it (S_{N-2}), and the second 74 dB tone (S_N) was compared to this biased memory. That is, assimilation occurs in memory. This accounts for why the judged ratio was large when S_{N-2} was small and small when S_{N-2} was large, and for why the response was often "1" when S_N and S_{N-1} were different (Lockhead & King, 1983, Figures 3, 5, 6, 7).

Magnitude estimation is a much more typical procedure for measuring psychophysical scaling functions than is this successive-ratios-judgment task. In magnitude-estimation tasks, subjects are instructed to judge the ratio between successive stimuli (just as above) and then to multiply that judgment by the previous response (Stevens, 1975). Thus, magnitude-estimation is a more complex task for the subjects than is our successive-ratios-judgment task. There is assimilation in magnitude-estimations (Ward, 1970) just as there is in successive-ratio-judgments and in absolute-identifications. Too, numerical responses are not needed for this to occur. There is assimilation in cross-modality matching data where people match the loudness of a tone to the duration of a key press (Ward, 1975).

This appears to be a ubiquitous result. There is assimilation in every set of psychophysical scaling data that has been examined and reported, no matter what the experimental procedure (DeCarlo & Cross, 1990; Luce & Green, 1978; Marks, 1989; Purks, Callahan, Braida, & Durlach, 1980; M. Treisman, 1984; Ward, 1973; others). Psychophysical judgments depend on prior events.

Psychophysical models and sequence effects. Traditional psychophysical scaling models (Equations 1-4) are based on average responses to each stimulus. Because there are sequence effects, this means that such models cannot reliably predict individual judgments. To make this point clear, suppose a magnitude estimation study in which the average of all responses to a 70 dB tone was 150 and the best fitting power function also indicated a response of 150 to that tone. Then, 150 is the best estimate available, from the scaling model, of the response to that tone on individual trials. However, 150 might never have been assigned to that tone. Such a result often occurs (Lockhead & King, 1983). This is because 150 is the average of smaller responses when the previous tone was quiet and larger responses when the previous tone was loud. Hence, actual responses are not predicted well by psychophysical models. They are predicted well when context is considered (Lockhead & King, 1983, Equation 1).

Just because there are context effects does not necessarily mean there is not an underlying psychophysical scale. An example from physics makes this obvious. When measuring the rate at which objects fall, wind makes it difficult to evaluate the gravitational constant, *g*. Nonetheless, as supported by measurements in a partial vacuum and by converging theories and data, the constant is real.

Similarly, factors that produce context effects in psychophysical judgments might make it difficult to measure a psychophysical scale which might also be real. Indeed, for purposes of psychophysical theory Luce and Krumhansl, among others, observed that effects of sequence are "often viewed as a mere nuisance" (1988, p. 52). This is because these effects interfere with the search for the underlying scale, perhaps as air interferes with measuring *g*. By this view, it is only essential to remove context effects to demonstrate the sought scale.

But the situation is not really this simple and Luce & Krumhansl are not as sanguine as their above quote may suggest. They summarize several demonstrations that the view is in difficulty and they closed their chapter in Stevens' Handbook of Experimental Psychology with the observation that "One cannot but be concerned by the demonstration (King & Lockhead, 1981) that the exponents [of psychophysical scaling functions] can easily be shifted by as much as a factor of 3 ... Clearly, much more work, using the data from individual subjects, is needed before we will be able to develop any clear picture of the structure of psychophysical scales." (1988, p. 67)

While more work surely needs to be done, I know of no evidence to suggest that new insights might come from studying individual subjects. The view pursued here is psychophysical scaling theory is difficult to demonstrate not because context makes testing difficult but because scaling theory is wrong.

II. Judgments in a Complex Situation.

Only univariate stimuli were considered above. Those limit the possibilities for demonstrating that it is relations among context and objects and attributes, not attributes themselves, that determine perceptions. A more complex stimulus situation provides this evidence more readily. One of the most compelling such demonstrations is the Ames distorting room (Ittleson, 1968). This is a room in which doors, windows, and floor tiles are (physically) trapezoidal rather than rectangular. The tall side of the trapezoid is further from the viewer than is the short side such that, when the room is viewed through a peephole, the trapezoids form right angles on the retina. There are no other depth cues to provide the correct information as to these shapes, and people perceive the room as a normal one with rectangular windows, doors, and floor tiles.

When a woman is viewed in such a room, her apparent size does not depend on how large she really is. It depends on where she is. She appears much larger in one corner than in another corner. This is because the corners are at different distances so she subtends a different visual angle when in a different corner, and is because the distances appear the same. According to the size-distance-invariance hypothesis, an object perceived to be at the same distance but with a different visual angle is perceived to be different in size, and so the woman is seen as small when, instead, she is far away.

This illusion of different sizes for identical objects in different environments does not require a complex room. Gregory (1970) showed that two equally tall people appear different in size when they are photographed at the same two distances as in the Ames room, but now without the room. In Gregory's demonstration, the people appeared in a photograph against a uniform white background with no depth cues present, except their feet were at the same elevation. They appear like a normal adult and a very small person standing side-by-side. This is because, again, the larger appearing (physically nearer) person subtends twice the visual angle of the other person, and because, which is different from the Ames room, the only cue to distance (ground position as indicated by the positions of their feet) is the people are at the same distance. This observation that objects appear different in size with or without the Ames room led Gregory to reject the Ames room as an experiment, because there is no control condition, and to conclude that "size difference is <u>not</u> attributed purely to distance" (p. 29).

Gregory's experiment is incomplete for his conclusion, and his conclusion is wrong. The correct control is again missing. This control is a view of people at different distances (they subtend different visual angles) in a real (ordinary) room which has ordinary texture cues (cf. Gibson, 1950). Then, perception is essentially veridical and the two people appear the same size.

Gregory actually made the same demonstration as Ames. Size judgment is determined by distance judgment, and thus size judgment is in error when distance information (context) is misleading. With correct context (the ordinary, structured world) size judgment for reasonably near events is essentially veridical. Rejecting Assumption III, it is the context and not the stimulus itself that determines judgments of attributes of the stimulus. This is the basis of object constancy. Constancy is provided by the context; objects themselves do not provide the information needed to judge them. <u>While it may be attractive to theorize about the processing of stimulus elements, it is necessary to theorize in terms of stimulus structures.</u>

Object constancy and the Ames room have been known for many years. However, the fact that context determines the perception of attributes has been persistently overlooked in attempts to demonstrate a true psychophysical function. This is an error. The complex settings of the ordinary world are not fundamentally different from the simple settings considered in the laboratory by psychophysical modelers. Judging the size of the woman (an attribute) is not essentially different from judging the size or brightness or some other attribute of a stimulus in magnitude estimation or absolute identification or other psychophysical experiments. Just as judgment of the height of a woman depends on her context, so does judgment of the brightness of a disc depend on its context. Indeed, the same disc appears dim or bright depending on what surrounds it. Again, we are unable to veridically abstract the magnitudes of attributes.

The situation is similar for sounds. The loudness of a tone depends on what other tones occurred. One difference from the brightness example is that the tones are presented sequentially and so the result is due to successive rather than simultaneous context. Thus, and now rejecting Assumption IV, judgments must be associated with memories of prior events. This memory involvement is not special to tones. Judgments of brightness and other attributes also depend on memory when stimuli are presented successively (Lockhead, 1970; in press).

III Judgments of Bivariate Stimuli.

Stevens (1975) noted that "all stimuli are multidimensional. Thus a simple patch of light presents many aspects" (p. 66), and people have "the ability to separate out of a complex configuration one single aspect and to compare that aspect with the same aspect abstracted from another configuration" (p. 66). This is Assumption II. This is examined directly in this section. People judged one attribute of a stimulus while another attribute varied from trial-to-trial. If the assumption is correct, then judgments of the relevant attribute should not be markedly affected by variations of the second attribute.

Three studies are reported. In the first, values of auditory loudness and pitch were correlated. The question asked is whether assimilation occurs to the individual attributes, as expected if attributes are judged separately, or if assimilation occurs to the complex stimulus, as expected if the entire stimulus is judged by comparing it to memories of other stimuli (Lockhead & King, 1983). The answer is the latter. In the second study, loudness and pitch were varied orthogonally and subjects judged the value of one or the other dimension. According to Assumption II, random variations in the irrelevant (not-judged) attribute do not matter because the relevant attribute is judged independently. The assumption is not supported. The third study examined the source of assimilation when loudness and pitch were both judged on each trial. The data are consistent with my *conclusion that assimilation occurs between memories of the* bivariate objects. The independence assumption is again not supported. Because these three studies have not previously been published, they are described in slightly more detail than the studies above.

<u>Context effects when stimulus attributes are correlated.</u> Judgments of bivariate stimuli were examined for sequence effects in absolute identification studies. The stimuli are indicated in the upper-left portion of Figure 4. These were ten auditory tones with loudness (amplitudes of 79 to 88 dB SPL in 1 dB steps) and pitch (frequencies of 1000 to 1045 Hz in 5 Hz steps) nonlinearly correlated. Amplitudes 1 through 10 were paired, consecutively, with frequencies 3, 6, 9, 1, 4, 7, 10, 2, 5, 8 (Lockhead, 1970, labeled these "sawtooth paired" as a mnemonic because connecting points along the X-axis produces a sawtooth-like figure). This produces a pairing of attributes across dimensions such that the amplitude of a stimulus perfectly predicts its frequency, and vice versa, although this correlation has a different form than the more commonly studied linear correlation.

--- Figure 4 about here ---

The subjects were told the structure of the stimulus set and were given a key to refer to whenever they so chose during the experiment. Four people were asked to identify only the intensity of each randomly presented tone when feedback (the numerals 1-10 correlated with intensity) was given and, separately, when feedback was not given after each response. There were 400 trials for each subject in each condition. Half of the subjects performed the feedback task first and half did the no-feedback task first. None of these subjects were ever asked to judge pitch.

Because the response and feedback numbers are both correlated with loudness, and because there is no uniform relation between pitch and loudness or pitch and response, the subjects might have ignored pitch and attended only to loudness. In that case, any sequential structure would be associated only with intensity. If the stimuli are integral such that people cannot attend to one attribute independent of variations in other attributes, but process the entire stimulus before abstracting an attribute value (Lockhead, 1972), then sequential structure might be associated with pitch as well as with loudness.

As a hypothetical example, consider those trials when stimulus #5 was presented. If the entire stimulus was initially judged, then it, both its pitch and its loudness, might assimilate toward the prior total stimulus, toward its pitch and its loudness (or their combination). In that case, responses to stimulus #5 would tend toward the value of the previous bivariate stimulus. The collection of all such responses would then reflect the structure of the bivariate stimulus space.

This is apparently what happened. The upper-right portion of Figure 4 shows the average response to stimulus #5 as a function of the response on trial N-1 when feedback was not given. These sequence effects reflect the structure of the stimulus space. When the prior response was #1, the response to stimulus 5 (amplitude 5, frequency 4) tended to be 1 or 2 or 4, as if it were identified as quiet and low pitch, i.e., toward the value of the prior trial. When R_{N-1} was #2, stimulus 5 tended to be identified as slightly louder (and higher in pitch) than when the prior response was #1. It was more often called 3 than 1. This tendency for responses to stimulus 5 to be similar to the value of the prior response in terms of its location in the X-Y domain, rather than only along the X domain, is seen for all ten sequences.

This same analysis was made for each of the ten stimuli. The structure of the averaged responses to each stimulus as a function of the prior response reflects the structure when 5 was the stimulus. This is seen in the ten outlined regions in the bottom portion of Figure 4. These enclose the responses to each stimulus. The numerals within each region are the responses that had been given on trial N-1. The position of each numeral indicates the median response, calculated by separately averaging X and Y coordinates of the response on trial N. Although there is variability, some of which may be due to the small amount of data and the fact that subjects did not use all responses equally often, the sequential structure in each of the ten response sets reflects the distribution of the parent set of ten stimuli.

The above analysis is for data collected when there was no feedback. When feedback was given after each response, there again was assimilation to both the prior stimulus (or feedback) and the prior response. The difference compared to the no-feedback data is assimilation was greater between successive stimuli than between successive responses, whereas, when feedback was not given assimilation was greater to prior responses than to prior stimuli.

Response times also depended on sequence. Figure 5 shows the median response times to identify each stimulus, as a function of the distance between it and the previous stimulus, when feedback was given. Responses were faster when successive stimuli were more similar [with similarity measured as the Euclidian distance between stimuli in the frequency-amplitude space] ($r = 0.87$). The magnitude of the effect is large. It is about 800 ms when stimulus repetitions are included and about 600 ms when repetitions are not included. Consistent with this, response times also correlate with the difference between successive responses ($r = 0.68$).

In the no-feedback data (not shown), response times again correlate with the difference between successive stimuli ($r = 0.68$) and with the difference between successive responses ($r = 0.72$; all ps < 0.01).

---Figure 5---

Conclusion. Although only one attribute was to be judged, loudness, responses assimilated toward both attributes of the prior stimulus, toward both pitch and loudness. This is consistent with the suggestions here that each stimulus is perceived in terms of the memory of the prior total stimulus, and there is assimilation between successive events. Only after that processing did subjects judge the loudness of the current stimulus, which they did in terms of the pitch as well as the loudness of the prior tone. I conclude that successive, integral stimuli are compared in memory, there is assimilation between stimuli in that multidimensional space, and attribute judgments are based on an analysis of the assimilated, total stimulus.

This conclusion, which is based on averaged judgments or classifications, is consistent with the fact that responses took longer on trials that successive stimuli were more different from one another in the bivariate space. It is as if more time is required to evaluate the relation between stimuli that are more distant from one another in the metaphorical memory or similarity space (Hutchinson & Lockhead, 1977; Monahan & Lockhead, 1977; Lockhead, in press).

These results with bivariate stimuli extend the previously summarized findings with univariate stimuli. In both classes of data: 1) Assimilation is greater to the prior response than to the prior stimulus when there is no feedback, 2) assimilation is greater to the prior stimulus than to the prior response when there is feedback, 3) there is always assimilation, and 4) responses take longer when successive stimuli are more different. Because judgments depend on other attributes in the stimulus and depend on sequence or time, Assumptions II and IV are rejected.

Context effects when stimulus attributes are orthogonally paired. In the above study, people judged one attribute when another attribute was correlated with it. Thus, both attributes might have been expected to be attended by the subjects, which they were. It cannot be decided on the basis of only those data whether or not people are able to attend to one attribute and avoid others. They might have attended to both attributes because both were informative for identifying the total stimulus.

In this experiment, people were asked to judge the value of one dimension when the value of second dimension varied randomly from trial to trial. Here, the second, dimension contains no useful information. It is irrelevant to the task and might sensibly be ignored.

The dimensions were again auditory amplitude and frequency. People judged the relevant dimension (loudness or pitch) while the irrelevant dimension (pitch or loudness) varied from trial-to-trial by a lot, or by a little, or not at all.

Specifically, when loudness was judged, all tones were 70 dB or 72 dB loud and presented randomly. In three experimental conditions, these intensities were presented at, also randomly selected, 1000 and 1015 Hz (narrow range), or 1000 and 1045 Hz (intermediate range), or 1000 and 1500 Hz (wide range). In two control (univariate) conditions, these intensities were presented at fixed frequencies, 1000 Hz or 1500 Hz. Fletcher & Munson's data show that loudness is virtually independent of frequency at these levels (Stevens & Davis, 1938, p. 123 ff).

Analogously, when pitch was judged the randomly selected frequencies were always 1000 or 1015 Hz. In four experimental conditions, these frequencies were presented at, also randomly selected, 70 and 72 dB (2-spread), 70 and 76 dB (6-spread), 70 and 80 dB (10 spread), or 61 and 91 dB (30 spread). In two control conditions, these two frequencies were randomly presented at 61 dB or at 80 dB.

Six subjects each gave 400 responses to each condition. When loudness was judged the subjects classified each tone as quiet or loud, and when pitch was judged they classified each tone as low or high, by pressing the left or right of two buttons as quickly and accurately as they could. Each tone was presented for 200 msec. There was 500 msec between the response and the next tone. No feedback was given.

<u>Results.</u> For all noted results, $p < 0.01$. For every comparison available, performance was faster in the univariate (control) conditions than the orthogonal (experimental) conditions. In all cases, errors correlated positively with response times.

Median response times when loudness was judged are shown in Figure 6 as a function of whether or not frequency repeated on successive trials. Overall, median responses in the wide range condition (frequency variations of 500 Hz) were slower (765 ms.) than in the intermediate and narrow range conditions (variations of 45 Hz and 15 Hz; RT = 618 and 600 ms and not reliably different), and responses in both of these conditions were slower than the univariate average (507 ms). This is an average range effect of 258 ms.

---Figure 6---

The data show this same form for when pitch was judged. Responses were fastest when intensity did not vary between trials (median RT = 508 ms), slower when intensity varied by 2 (576 ms), 6 (561 ms), or 10 dB (542 ms) [all of which were statistically equivalent], and slowest of all when intensity varied by 30 dB (605 ms). This is an average range effect of 97 ms.

Again, performance further depended on whether or not intensity repeated between trials. Responses were faster when that irrelevant attribute repeated than when it changed between trials.

All of these conclusions hold separately for when the level of the relevant stimulus repeated (response repetition) and for it changed between trials (response change).

Discussion. For stimuli that may vary between trials in both loudness and pitch, responses to classify levels of one dimension take longer, are more variable, and have more errors when the second dimension varies randomly from trial to trial than when the second dimension does not vary. Furthermore, the magnitudes of these effects are greater when the irrelevant dimension varies by greater amounts; the slopes in Figure 6 are positive. Assumption II is again rejected. It is not true that subjects judge one attribute independent of other attributes in the stimulus.

These results are consistent with earlier reports that variations in pitch affect loudness judgments and variations in loudness affect pitch judgments (Wood, 1973; Kemler-Nelson & Smith, 1979; Melara & Marks, 1990), except those studies did not manipulate the ranges of the irrelevant dimensions. These results also extend Felfoldy's (1974) demonstration [he used rectangles that varied in height and width as stimuli] that response times to classify bivariate stimuli depend on sequence.

The magnitudes of these effects depend on the amount by which the "irrelevant" dimension varies. Thus, the impairment of performance that is due to variations of an irrelevant attribute is not just a nominal effect. It is a quantitative effect with its magnitude correlated with the amount of trial to trial change in attribute values. Not only does judgment depend on other stimuli in the situation and their sequencing, but the magnitude of these effects depends on the magnitudes of these contextual differences. This is consistent with the view, already offered, that successive stimuli are compared and this comparison is easier to make when the stimuli are more similar to one another in memory.

Assimilation of bivariate stimuli occurs in memory. It was demonstrated above that assimilation occurs in judgments of bivariate stimuli, just as in judgments of univariate stimuli. In discussing univariate judgments, I concluded that assimilation occurs in memory [as well as between successive responses]. To learn if assimilation in bivariate judgments also occurs in memory, I asked four people to identify both the loudness and pitch of ten tones.

The stimuli were the ten sawtooth paired auditory amplitudes and frequencies used in the identification study reported earlier and described by the upper left portion of Figure 4. The subjects in that identification study knew there were ten tones and knew the structure of the correlation between the dimensions. Here, the subjects were told nothing about the stimulus set. They were simply asked to categorize each loudness with the numbers 1-10 and, on each trial, to then categorize each pitch with the numbers 1-10.

Various outcomes might be predicted. 1) If subjects come to learn the stimuli after some number of trials, then they will know that only ten different tones are involved. 2) If assimilation occurs between successive perceptions, then the subjects might hear each tone as overly like each previous tone. In this case, the various tones might appear very similar and the subjects might conclude that few stimuli are involved. 3) If assimilation occurs between memories of successive stimuli, such that each stimulus is compared to the memory of the prior stimulus (which had assimilated toward the memory of the tone before it), then there could be 10 X 10 memories (or more), one for each stimulus that assimilated in memory toward each already assimilated memory.

Results and discussion. After 400 trials, each subject was asked unexpectedly to estimate how many different tones had been presented. These estimates by the four subjects were 68, 80, 98 and 100 different tones. Following an additional session of 400 trials on the next day, these estimates were 37, 50, 75, and 90 tones. This last outcome is despite the possibility that the subjects were alerted in the prior session that something about the number of tones was important. Of course, only the same ten tones had been presented over the entire 800 trials.

According to each subject's report when asked at the end of the study, the actual correlation between pitch and loudness was never detected. Too, each subject used 90 or more of the possible 100 responses (each of the 10 loudness responses X each of the 10 pitch responses) during the study.

These guesses by the subjects as to the number of stimuli in the study are consistent with the earlier conclusion that assimilation between successive tones occurs in memory. Because each tone is preceded by all tones, at least 100 memories of the 10 tones were available. These results are also consistent with the memory model offered by Lockhead and King (1983; also see Holland's, 1968, regression model; M. Treisman's, 1984, criterion-setting model; DeCarlo & Cross's, 1990, regression model; and Killeen's, 1989, statistical search model which I find particularly attractive) who proposed that the response, R_N, assimilates toward the memory of the stimulus on the previous trial and contrasts from memories of earlier trials:

$$R_N = S_N + a(M_{N-1} - S_N) + b(\bar{M} - M_P) \qquad [5]$$

where S_N is the stimulus, M_{N-1} the memory of the previous stimulus, $\bar{M}$ is the average memory of all stimuli during the experiment, M_P is the average memory of stimuli on trials N-2 to N-7 and called the memory pool, and a and b are positive constants.

The data reported here allow two extensions to the model expressed by Equation [5]. The first is that the model, which was based on univariate data, generalizes to multivariate data. The second is that assimilation occurs between memories of objects in the complete memory space, rather than just along responses or the judged dimension.

Conclusion

At least sometimes, it is wrong to assume that attributes are judged independently of other attributes in a stimulus. Much evidence supports this conclusion. Hue is determined by relations among wavelengths (Land, 1959), apparent shape depends on total stimulus structure (Ittleson, 1968; many perception demonstrations), brightness is decided by contrasts over time and space (Arend et. al, 1971; Cornsweet, 1970), and apparent size depends on apparent distance (Emmert's law) which, in turn, depends on structural relations in the environment (Gibson, 1950, 1979; Lockhead & Wolbarsht, 1989). Each such fact demonstrates that physical values of attributes not only do not determine performance by themselves, they frequently are not even available to perception.

Because of such effects, because there are so many different ones, because some of them interact, and because judgments also vary from trial-to-trial due to task and sequence differences, any underlying, true psychophysical scale can only appear in the data as a will-o'-the-wisp with no basis to decide which observed scale is the "true" scale. Except for its esthetic appeal, which is considerable, there seems to be little reason to expect a fixed relation between behavior and the amount of energy in some attribute of a stimulus, and little reason to expect to be able to demonstrate such a function should one exist.

This is really not news. Many of the difficulties noted here have been known for a long time. However, unlike physics, psychophysical models have not yet given way to another view. One reason is scaling models have practical value. A bril scale is convenient for rating light fixtures and a sone scale is useful for designing music halls. This, alone, is sufficient reason to continue and even expand use of such scales. Another reason the general model has not been replaced may be that an equation that correlates as well with phenomenology as does the power function is satisfying in itself and no need for a different level of explanation is compelling. Still another reason may be that it is so difficult to understand everything involved when we judge objects that it is attractive to "Keep it simple" (Krueger, 1989, p. 311).

But perception and judgment are not simple. Instead, the evidence warrants searching for a theory based on something other than physical intensity and phenomenology. Indeed, why might one expect there to be a true psychophysical scale? What would it mean? Must it reflect a psychophysical parallelism, or dualism, or other separation of physical and psychological worlds? A reviewer of an earlier version of this paper thought so: "I think it [a psychophysical world] means to affirm the existence of a world of psychical reality above and beyond the neurophysical basis for perception."

This reviewer summarized the manuscript as follows: "(a) Under constrained conditions it is possible to get regular functional relations between physical entities such as sound intensity or luminance and verbal or other judgments. (b) These relations are termed psychophysical laws for reasons having their origin in Fechner's eccentric conflation of physics and psychology. (c) But these relations are highly labile: they depend on context; they do not reflect highly reliable sequential effects; they cannot be obtained at all unless certain preconditions (such as movement for visual stimuli), not an explicit part of the law, are met; they assume an independence of stimulus attributes for which the evidence is almost all negative. (d) Thus, the case is overwhelming that the psychophysical laws, though regular and reproducible, are largely irrelevant to the principles of operation of the behaving organism -- which is what psychology is presumably all about. Perhaps any mechanism that is capable of the feats of pattern recognition of which humans are capable would show similar laws -- and they would be similarly unrelated to the details of its operation." [2]

Might A biological view be more productive?

Perhaps a search for understanding psychophysical judgments should be based on the biological sciences rather than on what is usually selected from the physical sciences to support Fechnerian psychology. This is because psychophysics is a study of reactions of complex biological organisms which evolved more to perceive things and events than to abstract and measure attributes. From a biological perspective, it is difficult to argue that knowing the intensity of an attribute of an object, such as its brightness, is fundamental. Such information is not only unessential for identifying objects in the natural world, it is often incompatible. The same object must be identified veridically in different environments where it may have different intensities.

This suggestion is not novel. In discussing mechanisms that detect intensity differences and in which absolute intensity levels are irrelevant or are lost, Cornsweet (1970, p. 379-380) said "it seems quite reasonable that information about relative intensities is more important for human survival than information about absolute intensities. It does not matter very much to a human what the absolute light level is (unless he is a photographer, and then he needs a light meter to regain the information his visual system has lost), but it is important for him to distinguish among different objects. It is also convenient that, with constancy, our perceptions are correlated with a property of objects themselves (i.e., their reflectances) rather than with the incident illumination."

Nonetheless, a biological perspective is seldom taken in writings about psychophysical scales. There are exceptions. Shepard (1987) proposed that sensory systems and mental representations evolved in terms of what is consequential to the animal. Thus, Shepard might seek generalization scales that mirror the animal's history. This thesis is different from his (1981) and Krantz's (1972) relational theory already mentioned. There, as in other psychophysical models, it is assumed that attribute intensity is the subject's stimulus and physical intensities are assigned numbers.

Another exception to the independence assumption is Warren's (1981) physical correlates theory where sensory scales are "linked to a view of perception as a dynamic system continually calibrating neurophysiological response to events and relationships present in the environment" (p. 189). It seems to me that this aspect of Warren's theory must be correct. However, Warren further assumes that the "physical correlate theory of sensory intensity leads directly to the rule that equal stimulus ratios produce equal subjective ratios" (p. 175). This does not necessarily follow from his correlates theory, although the two are consistent, and this is not supported by the evidence. It is rare to find data where equal stimulus ratios actually do result in equal responses ratios (Lockhead & King, 1983).

Such exceptions notwithstanding, most models of psychophysical scaling do not explicitly consider the biology of the subjects. Many examples are seen in the commentaries on Krueger's (1989) paper in this journal. While these 31 papers note difficulties with one or another theory, observe the inconvenience of particular data, argue that some model does not satisfy some observation, and observe how discouraging the search for a model has been, most (but not all) are firmly based in classical physics. Fechner's insight continues to be dominant.

Nonetheless, psycho-physical scaling models that mimic physical-physical models in the ways current ones do are wrong. This is important to note because there are costs associated with continuing to pursue them. Formulae with a compellingly simple form like R = f(I), particularly when presented with the word "Law" in a discipline where laws are rare, can easily be believed. This means they can also be misleading, and not only to psychophysicists. They can mislead anatomists and physiologists as to what to look for in attempts to understand the structural and functional bases of perception. They can also mislead cognitive and perception psychologists as to the processes involved in receiving, remembering, and responding to a stimulus and as to what models to build. And they can mislead engineers as to how to build an optimum environment.

What seems needed is a psychological or psycho-biological model of the processes that allow people and other animals to reliably identify things in a world of changing intensities and circumstances. This may have to be a complex model since different rules hold for different stimuli. Studies using integral stimuli (Lockhead, 1966, 1972; in press; Garner, 1974; Shepard, 1991) support the idea that objects are processed holistically before analysis of their components occurs. The stimuli in most psychophysical scaling studies are integral. However, studies that use separable stimuli are consistent with the opposite conclusion; attributes are processed first and identification of the total stimulus occurs only later (Treisman, 1986, 1990).

Stimuli are classified as integral or as separable depending on the outcomes of performance measures. Their attributes are integral if there is a redundancy gain when people classify correlated attributes and if there is interference when those attributes are orthogonally related. That is, performance in judging one attribute depends intimately on some other attribute. Stimuli are separable when this is not the case, when variations in one attribute do not affect performance on another attribute. A reason for these differences is not known.

Although binary classifications like this almost always turn out to be an oversimplification, it might be useful to enquire if this apparent difference between stimulus classes, integral and separable, is associated with differences between natural objects and manufactured objects. Features are essential to most functional, manmade objects. A chair without a place to sit is not a chair. Accordingly, Barton and Komatsu (1989) suggest that features may be defining for artifacts. This may not be the case for most naturally occurring objects. Putnam (1975) noted that a man without legs is still a man. In agreement with these views, Barton and Komatsu (1989) reported data consistent with the idea that natural objects are judged in terms of their chromosomal/molecular features (essences?) while artifacts are judged in terms of their functional features. Concerning the current paper, their suggestion allows the conjecture that natural objects tend to be integral while manufactured objects tend to be separable. If so, that would suggest that the noted performance differences are related to function and essence, another dichotomy.

A different speculation to account for the integrality-separability distinction concerns anatomy. Livingstone & Hubel (1987) described psychophysical data suggesting stimuli are analyzed by attributes in terms of magno-cellular and parvo-cellular neural systems (yet another dichotomy), where judgments depend on which attributes are processed by the same or by different neural systems. Their data allow the suggestion that processing might be integral when the relevant attributes are processed by one or by the other such system, whereas processing might be separable when different attributes are processed by different systems.

For now, these are only guesses (or wishes) to be entertained while debate concerning the classic dichotomy, that of elements versus wholes (Kubovy, 1985), continues.

General Conclusion. The formulation R = f(I) invites the inference that organisms react to amounts of attributes. This might be correct for some sensory systems but it is not true for perceptual systems. A primary task for perception is to evaluate things in the environment. To accomplish this, the system needs relations between stimulations. This is because the same object may stimulate sensory organs differently in different environments, and invariance across environments is needed to perceive the object as the same in each instance. For this purpose, evolution has apparently discovered the value of differences as opposed to amounts. This is a key to the perceptual constancies and could be a basis for all of perception.

There is no reason to suppose that perception functions differently in psychophysical experiments than in the ordinary world. The essential stimulus in both cases is the collection of sensed differences across space and time. This is what needs to be modeled. For this purpose, it might be appealing to address this general problem of defining the stimulus in terms of parallel distributed processing (PDP) models (Rumelhart et al., 1986). Those allow examining effects of many factors at one time. In that case, a caution noted here must also be attended there. This is that, much like univariate scaling models, most PDP models are also based on features and magnitudes rather than differences and relations (this is not necessary; see Grossberg, 1976; 1988) and differences are the key to understanding perception.

A comment on method

Science is largely a search for invariance and for its explanation in terms of mechanisms. During our search, we sometimes discover regularities. Regularity and invariance are not the same thing. Regularity is to be expected when procedures are repeated. Invariance requires regularity across procedures and situations.

Data that are regular sometimes have an appealing form as well as repeatability. A salient example here is when responses are linearly related to stimuli along experiment r selected scales. This can be alluring. However, such linearity does not imply a mechanism. It may not even reflect one. Converging studies are required to demonstrate the validity of explicit and implicit assumptions concerning the sources of the noted structure in the data.

This is well known and probably everyone would agree. However, this can be easy to forget when the linear data are generated by the individual experimenter. This gives the added attraction of personal experience. This also occurs often in psychophysics. Many of us first explored, say, the method of magnitude estimation as graduate students. Most of us used essentially the same procedure and got essentially the same results. We used stimuli that ranged from near threshold to near pain and that were spaced equally on a logarithmic scale. The lights or sounds, usually, were presented in random order against a dark or quiet background and were judged in regard to a modulus of 100 for the middle value or to no modulus. Our data usually fit a power function pretty well. At this level, Stevens', Fechner's, and related equations are true.

However, the equations are not general. Changing the procedure changes the outcome. It varies with stimulus set (Garner, 1954), stimulus sequence (Holland & Lockhead, 1968; Luce & Green, 1978, Lockhead & King, 1988), stimulus range (Gravetter & Lockhead, 1973; Teghtsoonian, 1971), stimulus spacing (Parducci & Perrett, 1971; Weber, Green, & Luce, 1977), background (Brysbaert & d'Ydewalle, 1989), information feedback (Ward & Lockhead, 1971), and other factors. Because psychophysical scaling data are not invariant across procedures, psychophysical scaling models may describe particular data sets but nothing more.

Too, scaling models do not reflect the mechanisms that produce the data. Sequence effects demonstrate that judgments depend on prior events. However, scaling models average the data across sequences and thus at least partially obscure the mechanisms responsible for them. Such difficulties occur, in part, because psycho-physics has been based on physical models with assumptions that are not appropriate to psychology. The physical intensities of attributes are not directly available to perception and are not of primary relevance to the organism. In physics, whether gold or feathers are placed on a balance pan, weight can be selected out and measured (judged) independent of other attributes or of what was weighed previously. This is not true for organisms. They do not judge attributes independent of context. Rather than being an aberration, the size-weight illusion reveals the ordinary functioning of perceptual systems.

One advantage of the proposed biological perspective over a physical one is biology does not have these assumptions. However, this does not mean biology is the answer. It has its own assumptions and rules and we cannot know <u>a priori</u> what ones are appropriate for psychology. Rather, we must persistently examine the fundamental although often unstated assumptions of <u>any</u> theory that is promoted.

Once a regularity has been found, a common approach in psychology is to suggest an explanatory theory and seek additional data consistent with that theory. That should continue. But, particularly if that process has been highly successful, we must also seek data that are inconsistent with the theory, data that provide boundary conditions for the theory, and data that test its foundations. It is difficult to undertake such research when it questions firmly held beliefs but it is essential.

Footnote.

[1]Supported by AFOSR87-0353 to Duke University. I thank the Air Force Office of Scientific Research for their generous support, Herbert Crovitz, Peter Holland, Lester Krueger, Donald Laming, R. Duncan Luce, Larry Marks, Fritz Müller, John Staddon, and three anonymous reviewers for their comments, and David Rubin for many valuable suggestions.

[2] Anonymous.

References

Aiba, T.S. & Stevens, S.S. (1964) Relation of brightness to duration and luminance under light- and dark-adaptatio . Vision Research, 4, 391-401.

Algom, D. & Marks, L.E. (1990) Range and regression, loudness scales, and loudness processing: Toward a context-bound psychophysics. Journal of Experimental Psychology: Human Perception and Performance, 16, 706-727.

Allik, J. (1989) Is a unified psychophysical law realistic? Brain and Behavioral Sciences, 12, 267-268.

Anderson, N.H. (1981) Foundations of Information Integration Theory. New York: Academic Press.

Anglin, J.M. & Mansfield, R.J.W. (1968) On the brightness of short and long flashes. Perception & Psychophysics, 4, 161-162.

Arend, L. E. (1970) Contour Factors in the Broca-Sulzer Phenomenon. Ph.D. dissertation, Duke University.

Arend, L.E., Buehler, N. & Lockhead, G.R. (1971) Difference information in brightness perception. Perception & Psychophysics, 9, 367-370.

Arend, L.E. & Timberlake, G.T. (1986) What is psychophysically perfect image stabilization? Do perfectly stabilized images always disappear? Journal of the Optical Society of America, 3, 235-241.

Barton, M.E. & Komatsu, L.K. (1989) Defining features of natural kinds and artifacts. Journal of Psycholinguistic Research, 18, 433-447.

Birnbaum, M.H. (1982) Controversies in Psychological Measurement. In Social attitudes and psychophysical measurement, Erlbaum.

Bloch, A.M. (1885) Experiences sur la vision. Comptes Rendus de la Societe de Biologie, 37, 493-495.

Bolanowski, S.J. & Gescheider, G.A. (1991) Ratio Scaling of Psychological Magnitude In Honor of the Memory of S. S. Stevens, Erlbaum, NY.

Boring, E.G. (1950) A History of Experimental Psychology. Prentice-Hall, NJ.

Broca, A. & Sulzer, D. (1902) La sensation lumineuse en fonction du temps. Journal de Physiologie, Paris, 4, 632-640.

Brysbaert, M. & d'Ydewalle, G. (1989) Unifying psychophysics: And what if things are not so simple? Brain and Behavioral Sciences, 12, 271-273.

Campbell, F. W. & Robson, J.G. (1961) A fresh approach to stabilized retinal images. Journal of Physiology, 158, 11-12P.

Cornsweet, T.N. (1970) Visual Perception, Academic Press, NY.

DeCarlo, L.T. & Cross, D.V. (1990) Sequential effects in magnitude scaling: Models and theory. Journal of Experimental Psychology: General, 119, 375-396.

Falmagne, J-C (1985) Elements of Psychophysical Theory. Oxford Psychology Series No. 6, Oxford University Press, N.Y. 320-322.

Felfoldy, G. L. (1974) Repetition effects in choice reaction time to multidimensional stimuli. Perception & Psychophysics, 15, 453-459.

Garner, W.R. (1954) Context Effects and the Validity of Loudness Scales. Journal of Experimental Psychology, 48, 218-224.

Garner, W.R. (1958) Advantages of the Discriminability Criterion for a Loudness Scale. Journal of the Acoustical Society of America, 30, 105-1012.

Garner, W.R. (1974) The processing of information and structure, Hillsdale, NJ: Lawrence Erlbaum.

Gescheider, G.A. (1988) Psychophysical scaling. Annual Review of Psychology, 39, 169-200.

Gibson, J.J. (1950) The Perception of the Visual World. Boston, Houghton Mifflin.

Gibson, J.J. (1979) The Ecological Approach to Visual Perception. Boston, Houghton Mifflin

Graham, C.H. & Kemp, E.H. (1938) Brightness discrimination as a function of the duration of the increment in intensity. Journal of General Physiology, 21, 635-650.

Gravetter, F. & Lockhead, G.R. (1973) Criterial range as a frame of reference for stimulus judgment. Psychological Review, 80, 203-216.

Gregory, R.L. (1970) The Intelligent Eye. McGraw-Hill, NY.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: Part I. Parallel development and coding of neural feature detectors. Biological Cybernetics, 23, 121-134.

Grossberg, S. (1988). Neural Networks and Natural Intelligence. Cambridge, MA: MIT Press.

Heinemann, E.G. (1950) Simultaneous brightness induction as a function of inducing and test-field luminances. Journal of Experimental Psychology, 50, 89-96.

Helson, H. (1964) Adaptation--level theory: An experimental and systematic approach to behavior. Harper, NY.

Hering, E. (1964) Outlines of a Theory of the Light Sense. (L. Hurvich, Trans.) Harvard Univ. Press, Cambridge, MA.

Holland, M. (1968) Channel Capacity and Sequential Effects: The influence of the immediate stimulus history on recognition performance. Ph.D. Dissertation, Duke University, Durham, NC.

Holland, M. & Lockhead, G. (1968) Sequential effects in absolute judgments of loudness. Perception & Psychophysics, 3, 409-414.

Hutchinson, J.W. & Lockhead, G. R. (1977) Similarity as distance: A structural principle for semantic memory. Journal of Experimental Psychology: Human Learning and Memory, 3, 660-678.

Ittleson, W.H. (1968) Size as a cue to distance, American Journal of Psychology, 64, 54-67; 188-202.

Katz, M.S. (1964) Brief flash brightness. Vision Research, 4, 361-373.

Kelly, D.H (1979) Motion and Vision II. Stabilized spatiotemporal threshold surface, Journal of the Optical Society of America, 69, 1340-1349.

Kemler Nelson, D. G. & Smith, L. B. (1979) Accessing similarity and dimensional relations: The effect of integrality and separability on the discovery of complex concepts. Journal of Experimental Psychology: General, 108, 133-150.

Killeen, P. (1989) Maximum entropy psychophysics. Read at 30th meeting of the Psychonomic Society, Atlanta, GA. Nov. 1989.

King, M.C. (1980) Sequential Effects in the Variability of Magnitude Estimations of Loudness, Ph.D. Dissertation, Duke University.

King, M. C. & Lockhead, G. (1981) Response scales and sequential effects in judgment, Perception & Psychophysics, 30, 599-603.

Krantz, D.H. (1972) A theory of magnitude estimation and cross-modality matching. Journal of Mathematical Psychology, 9, 168-199.

Krauskopf, J. (1963) Effect of retinal image stabilization on the appearance of heterochromatic targets. Journal of the Optical Society of America, 53, 741-744.

Krueger, L.E. (1989) Reconciling Fechner and Stevens: Toward a unified psychophysical law, Behavioral and Brain Sciences, 12,, 251-266.

Kubovy, M. (1985) Overview. In Handbook of Perception and Performance, Vol. 2, eds., K. Boff, L. Kaufman, and J. Thomas. Wiley, NY. Section VI.

Laming, D. (1985) Sensory Analysis, London, Academic Press.

Laming, D. (1988) Sensory Discrimination and Sensory Judgment: An Introduction to Modern Psychophysics, University of Cambridge, Department of Psychology.

Land, E.H. (1959) Color vision and the Natural image, Proceedings of the National Academy of Sciences, 45, 115-129 and 536-644.

Livingstone, M.S. & Hubel, D. H. (1987) Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. The Journal of Neuroscience, 7(11), 3416-3468.

Lockhead, G. R. (1966) Effects of dimensional redundancy on visual discrimination. Journal of Experimental Psychology, 72, 95-104.

Lockhead, G. R. (1970) Identification and the form of multidimensional discrimination space. Journal of Experimental Psychology, 85, 1-10.

Lockhead, G. R. (1972) Processing dimensional stimuli: A note. Psychological Review, 79, 410-419.

Lockhead, G. R. (1984) Sequential predictors of choice. In, Preparatory States and Processes, S. Kornblum and J. Requin (Eds.), Erlbaum, 27-48.

Lockhead, G. R. (1988) Modeling temporal and spatial differences. The Behavioral and Brain Sciences, 11, 302-303.

Lockhead, G. R. (in press) On Identifying Things: A case for context. B. Burns (Ed.) Percepts, concepts, and Categories: the Representation and Processing of Information. Advances in Psychology. North-Holland, Amsterdam.

Lockhead, G. R. & King, M. C. (1983) A memory model of sequential scaling tasks, Journal of Experimental Psychology: Human Perception and Performance, 9, 461-473.

Lockhead, G.R. & King, M.C. (1988) An alternative to the neural-attention hypothesis for auditory psychohysics. Fechner Day 88, 101-102, International Society for Psychophysics, Stirling, Scotland.

Lockhead, G.R. & Wolbarsht, M. L. (1989) The moon and other toys. In M. Hershenson (Ed.) The Moon Illusion, New Jersey: Erlbaum.

Luce, R. D. & Green D. M, (1978) Two tests of a neural attention hypothesis for auditory psychophysics. Perception & Psychophysics, 223, 363-371.

Luce, R. D. & Krumhansl, C. L. (1988) Measurement, Scaling, and Psychophysics. In R. Atkinson, R. Herrnstein, G. Lindzey, & R. D. Luce Stevens' Handbook of Experimental Psychology, (2nd Ed.), NY, Wiley Interscience, pp. 3-74.

Marks, L.E. (1974) Sensory proceses: The new psychophysics. NY, Academic Press.

Marks, L.E. (1989) On the principle of "Matching by Scaling". Paper presented at the 30th annual meeting of The Psychonomic Society, Atlanta GA.

Melara, R. D. & Marks, L.E. (1990) Dimensional Interactions in Language Processing: Investigating directions and levels of crosstalk, Journal of Experimental Psychology: Memory, Learning, and Cognition, 16, 539-554.

Monahan, J. S. & Lockhead, G.R. (1977) Identification of integral stimuli. Journal of Experimental Psychology: General, 106, 94-110.

Parducci, A., & Perrett, L.F. (1971). Category rating scales: Effects of relative spacing and frequency of stimulus values. Journal of Experimental Psychology, 89, 427-452.

Poulton, E.C. (1989) Bias in Quantifying Judgments. Erlbaum, Hillsdale, NJ.

Purks, S.R., Callahan, D.J., Braida, L.D, & Durlach, N.I. (1980). Intensity perception X. Effect of preceding stimulus on identification performance. Journal of the Acoustical Society of America, 67, 634-637.

Putnam, H. (1975) Is semantics possible? In H. Putnam (Ed.) Mind, language, and reality: Philosophical papers, Volume 2. Cambridge University Press, New York.

Sharpe, C. (1972) The visibility and fading of thin lines visualized by their controlled movement across the retina. Journal of Physiology, 222, 113.134.

Shepard, R. (1981) Psychological relations and psychophysical scales: On the status of "direct" psychophysical measurement. Journal of Mathematical Psychology, 24, 21-57.

Shepard, R. (1987) Toward a universal law of generalization for psychological science. Science, 237, 1317-1323.

Shepard, R. (1991) Integrality versus separability of stimulus dimensions: Evolution of the distinction and a proposed theoretical basis. In The Perception of Structure, G. R. Lockhead & J. Pomerantz (Eds.), American Psychological Association.

Staddon, J. E. R., King, M. C., & Lockhead, G. R. (1977) On sequential effects in absolute judgment experiments. Journal of Experimental Psychology: Human Perception and

Performance, 6, 290-301.
Stevens, J.C. & Hall, J.W. (1966) Brightness and loudness as functions of stimulus durations. Perception & Psychophysics, 1, 319-327.
Stevens, S.S. (1957) On the psychophysical law. The Psychological Review, 64, 153-181.
Stevens, S.S. (1961) To honor Fechner and repeal his law. Science, 133, 80-86.
Stevens, S.S. (1975) Psychophysics: Introduction to its perceptual, neural, and social prospects. G. Stevens (Ed.), NY, Wiley.
Stevens, S.S. & Davis, H. (1938) Hearing. J. Wiley, NY.
Teghtsoonian, R. (1971) On the exponents in Stevens' law and the constants in Ekman's law. Psychological Review, 78, 71-80.
Teghtsoonian, R. (1973) Range effects in psychophysical scaling and a revision of Stevens' law. American Journal of Psychology, 86, 3-27.
Treisman, A. (1986) Properties, parts, and objects. In Handbook of Perception and Performance, Vol. 2, eds., K. Boff, L. Kaufman, and J. Thomas. Wiley, NY. Chapt. 35.
Treisman, A. (1990) Features and objects in visual processing. In I. Rock (Ed.) The Perceptual World. Freeman, NY, 97-110.
Treisman, M. (1984) A theory of criterion setting: An alternative to the attention band and response ratio hypotheses in magnitude estimation and cross-modality matching, Journal of Experimental Psychology: General, 113, 443-463.
Ward. L. (1970) Some psychophysical properties of category judgments and magnitude estimations. Ph.D. dissertation, Duke University, Durham, NC.
Ward, L. (1973) Repeated magnitude estimations with a variable standard: Sequential effects and other properties. Perception & Psychophysics, 13., 193-200.
Ward, L. (1975) Sequential dependencies and response range in cross-modality matches of duration to loudness, Perception & Psychophysics, 18, 217-223.
Ward, L. & Lockhead, G. R. (1971) Response system processes in absolute judgment. Perception & Psychophysics, 9, 73-78.
Warren, R.M. (1981) Measurement of sensory intensity. The Behavioral and Brain Sciences, 4, 175-189.
Weber, D.L., Green, D.M., & Luce, R.D. (1977) Effects of practice and distribution of auditory signals on absolute identification. Perception & Psychophysics, 22, 223-231.
Wood, C. C. (1973) Levels of processing in speech perception: Neurophysiological and information-processing analyses. Ph.D. dissertation, Yale University.
Yarbus, A.L. (1967) Eye movements and vision, New York: Plenum Press.

Figure legends.

Figure 1. Top: Mean matching luminance as a function of the duration of a luminous disc for the three individuals who served in each duration condition (from Arend, 1970).

Bottom: Brightness (mean matching luminance) as a function of flash duration, in idealized form, for a low intensity (lower function) and a high intensity (upper function) light. (Adapted from Anglin & Mansfield, 1968.)

Figure 2 Rapidly spinning the spatial distribution mounted on a disc in the top panel produces the brightness distribution in the bottom panel. (Arend, Buehler, & Lockhead, 1971; photographs courtesy L. Arend)

Figure 3. The average effect of the stimulus on a given trial on responses to the next eight trials in an absolute judgment experiment; feedback was given after each response. Responses tend toward the stimulus value on trial N-1 (assimilation) and away from stimulus values on trials N-2 through about N-5 (contrast). (From Lockhead, 1984, with permission).

Figure 4. Upper left: The loudness-pitch (amplitude-frequency) pairings used as stimuli. Numerals in the figure are both the loudness levels and the stimulus-response labels assigned to each tone.

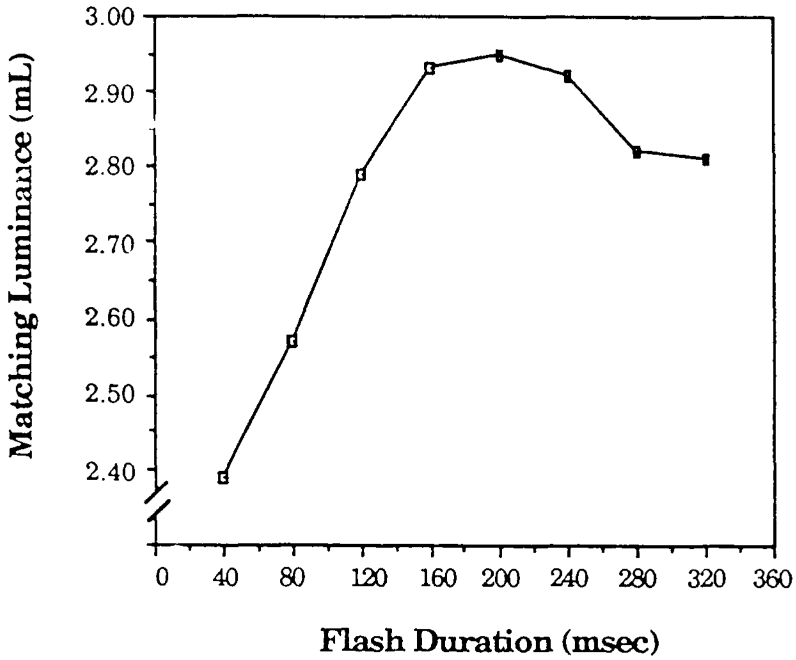
Upper right: Averaged responses to stimulus #5 (loudness value 5, pitch value 4 as in the upper left panel) as a function of the prior response. Each numeral is the response that was given on the prior trial. The location of each numeral is the median response, in X-Y coordinates, to stimulus #5 when it followed the noted response value.

Bottom: Averaged responses as a function of the prior response. This is the same as the upper right panel, except responses to all ten stimuli are now reported. Current stimuli are indicated by the ten outlinings, which have no other meaning. Within each outline, the numerals indicate the prior response value. The positions of the numerals indicate the median response to the current stimulus when it followed that prior response. There are two missing (superposed) data points due to identical response values.

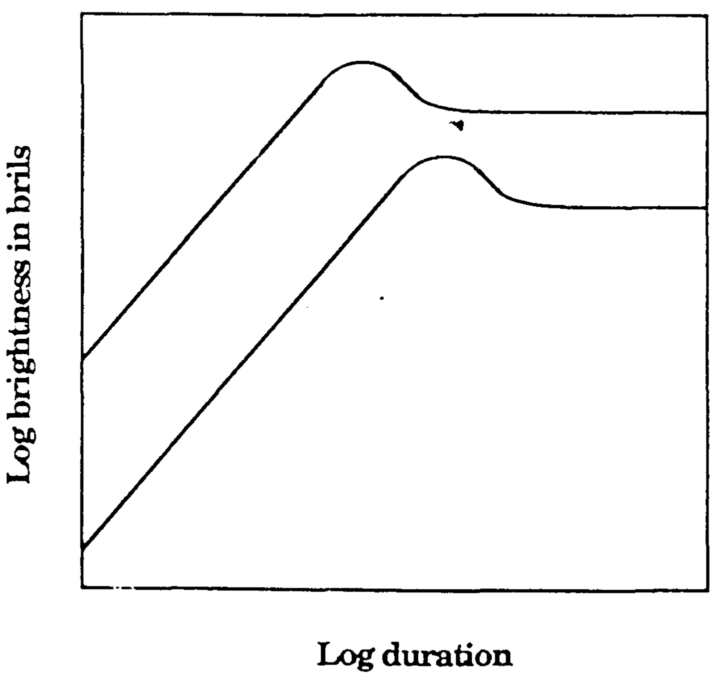
Figure 5. Median times to identify each stimulus as a function of distance, measured in the X-Y coordinate space of the upper left panel of Figure 4, between the current stimulus and the prior stimulus, when feedback was given.

Figure 6. Median response times to classify loudness on trials that pitch repeated (filled circles) and trials that pitch changed (open squares), as a function of the range of the irrelevant attribute in the orthogonal sorting tasks. The separated dots at 0 Hz range indicate response times in the two control conditions where loudness varied and frequency was always 1,000 or 1,500 Hz.

Figure 1



Matching Luminance (mL)
3.00
2.90
2.80
2.70
2.60
2.50
2.40
0
40
80
120
160
200
240
280
320
360
Flash Duration (msec)



Log brightness in brils
Log duration

Figure 2
(the Photo will reproduce better in print)

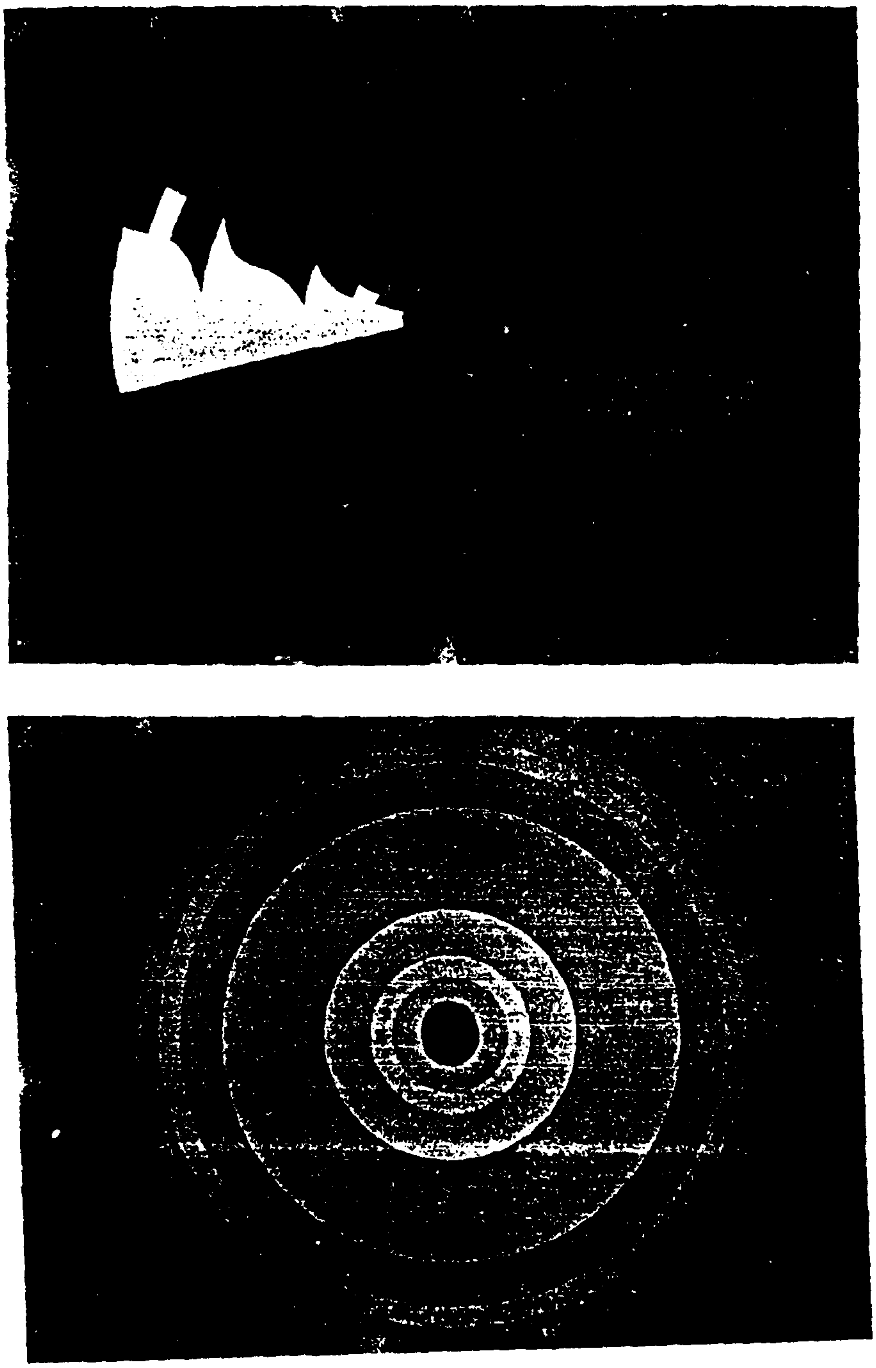
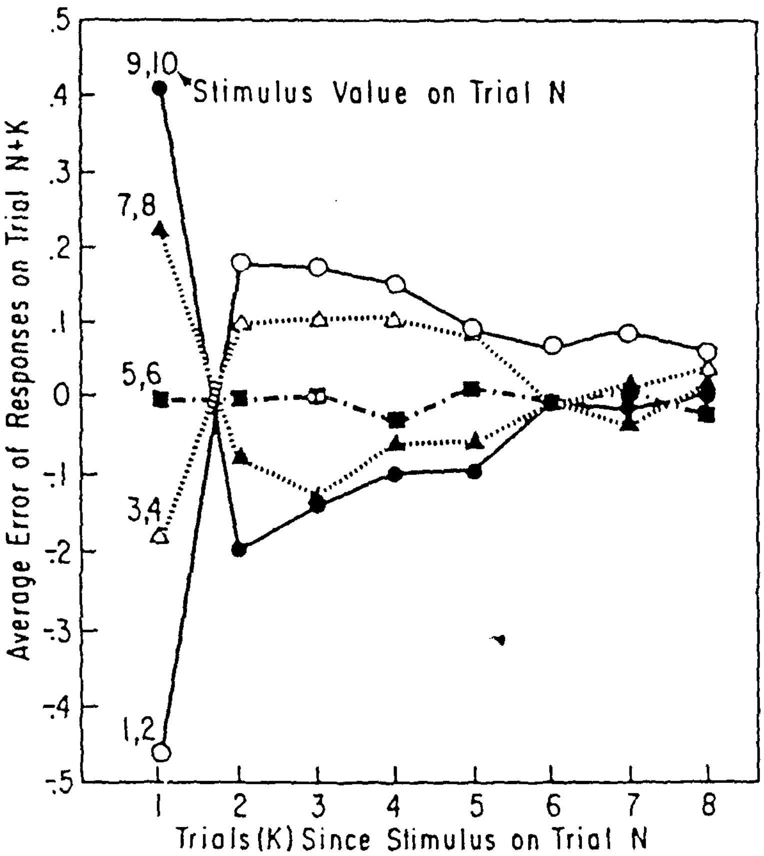
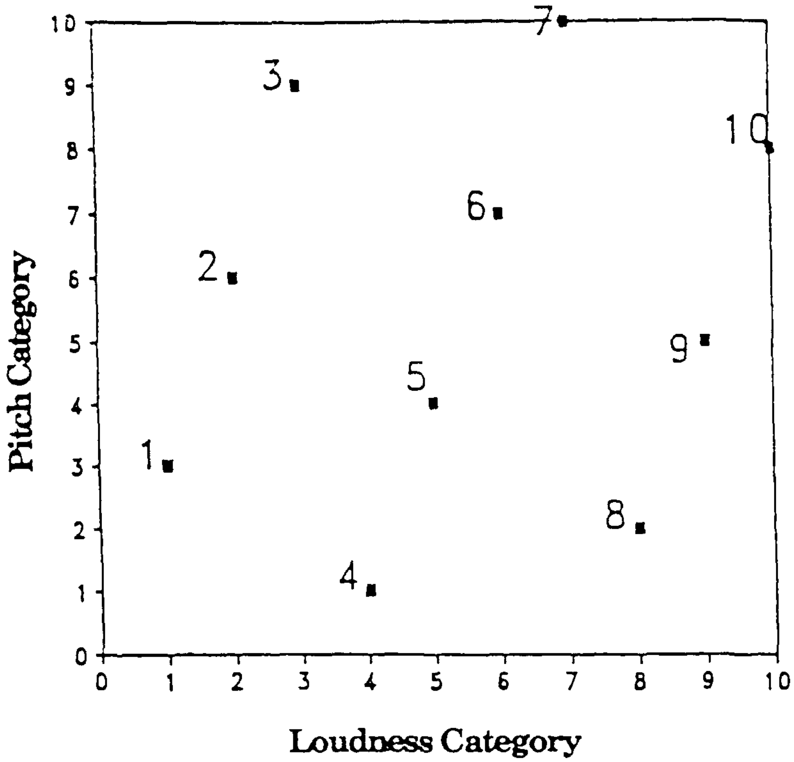


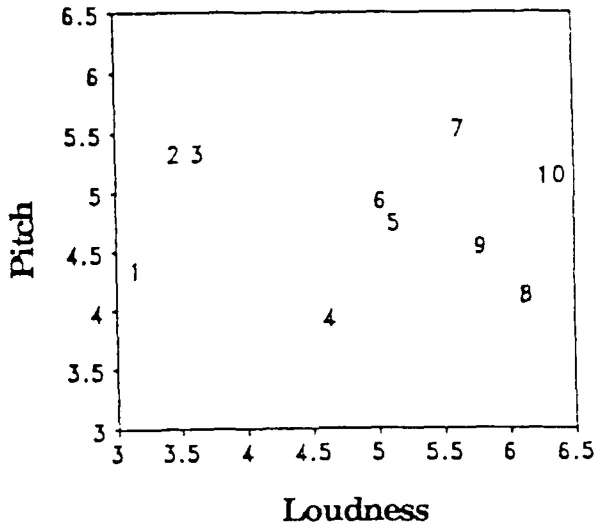
Figure 3



9,10
Stimulus Value on Trial N
7,8
5,6
3,4
1,2
.5
.4
.3
.2
.1
0
-.1
-.2
-.3
-.4
-.5
Average Error of Responses on Trial N+K
1
2
3
4
5
6
7
8
Trials (K) Since Stimulus on Trial N

Figure 4


Pitch Category
Loudness Category


Pitch
Loudness

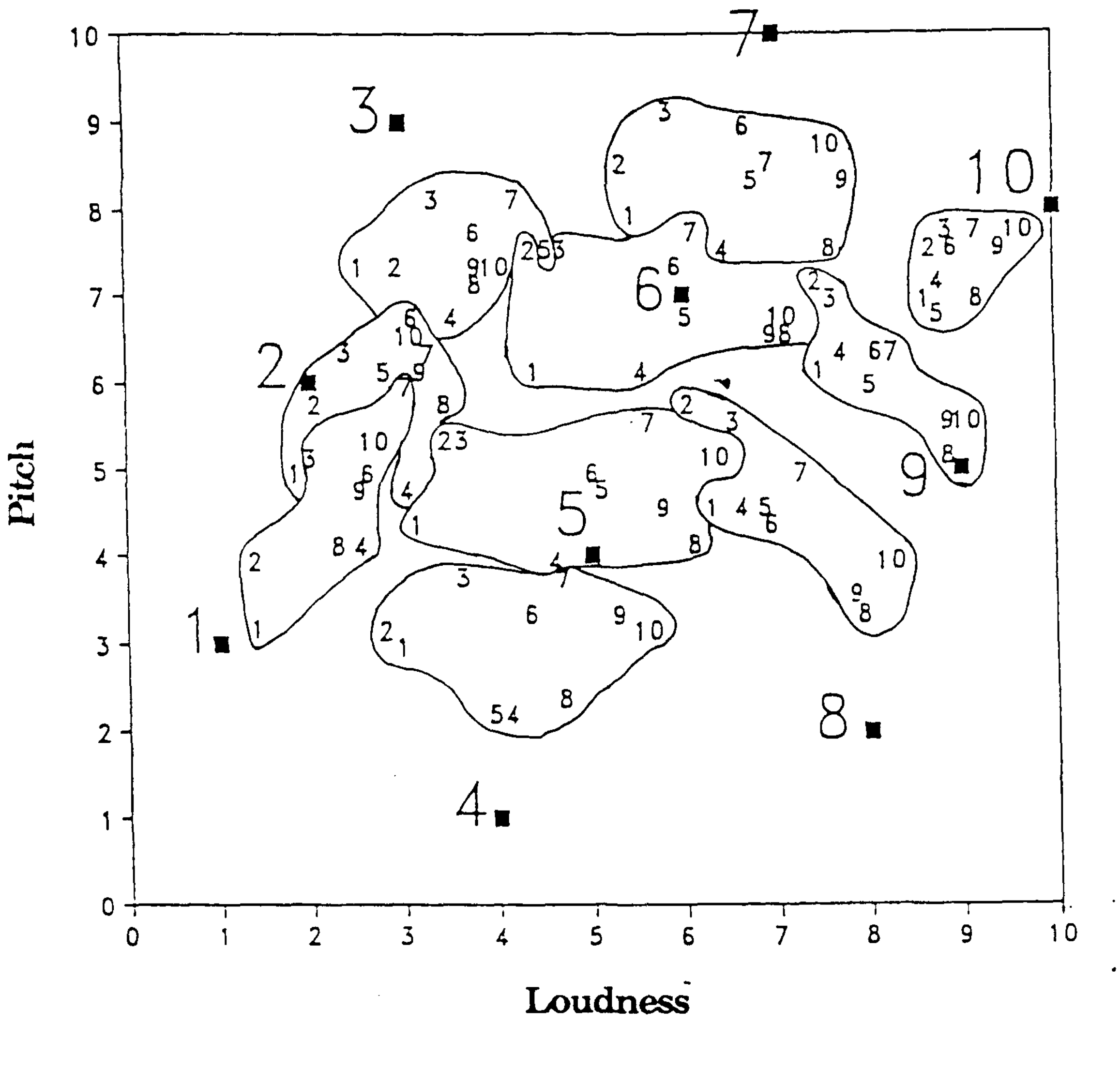

Pitch
Loudness

Figure 5

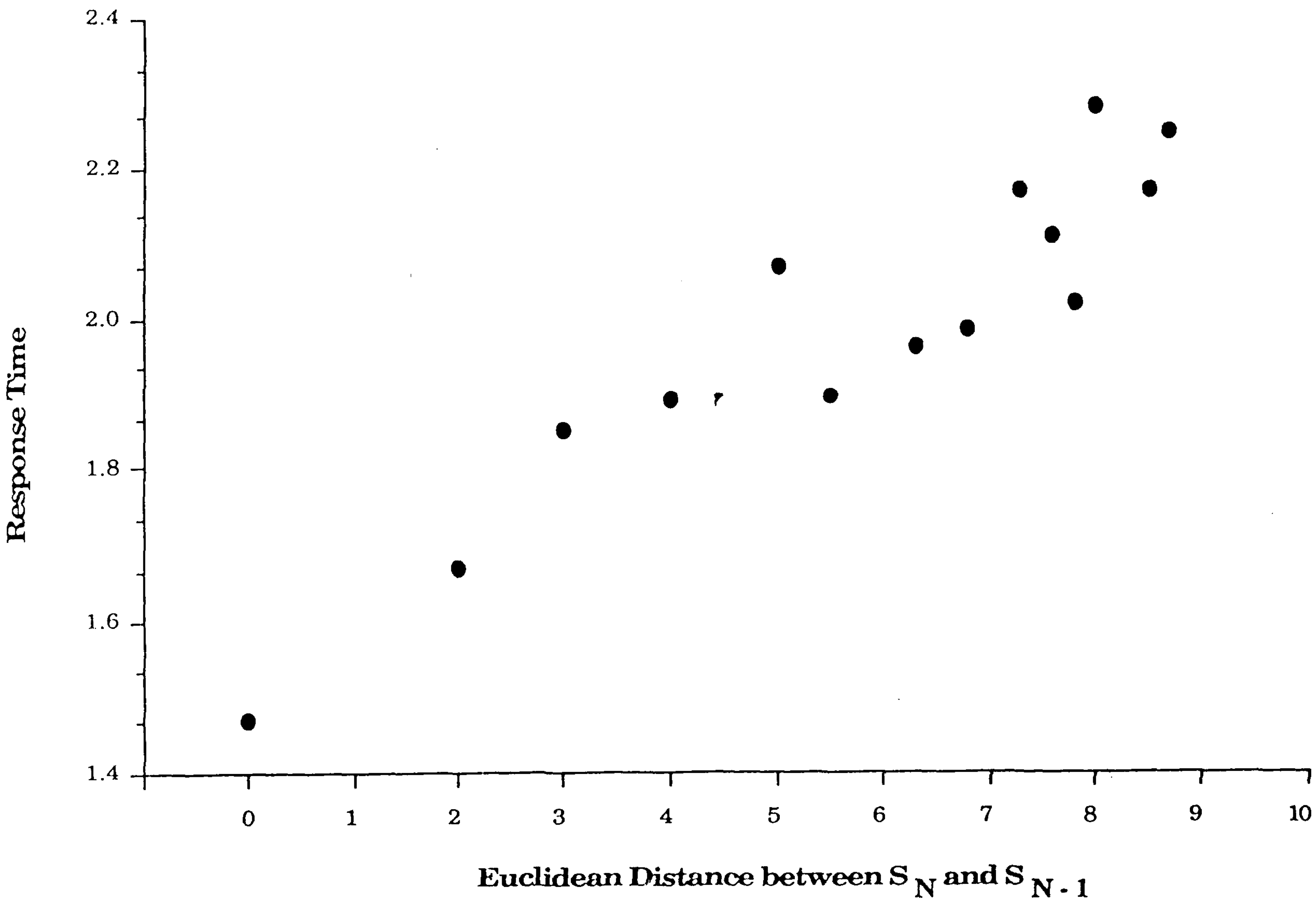
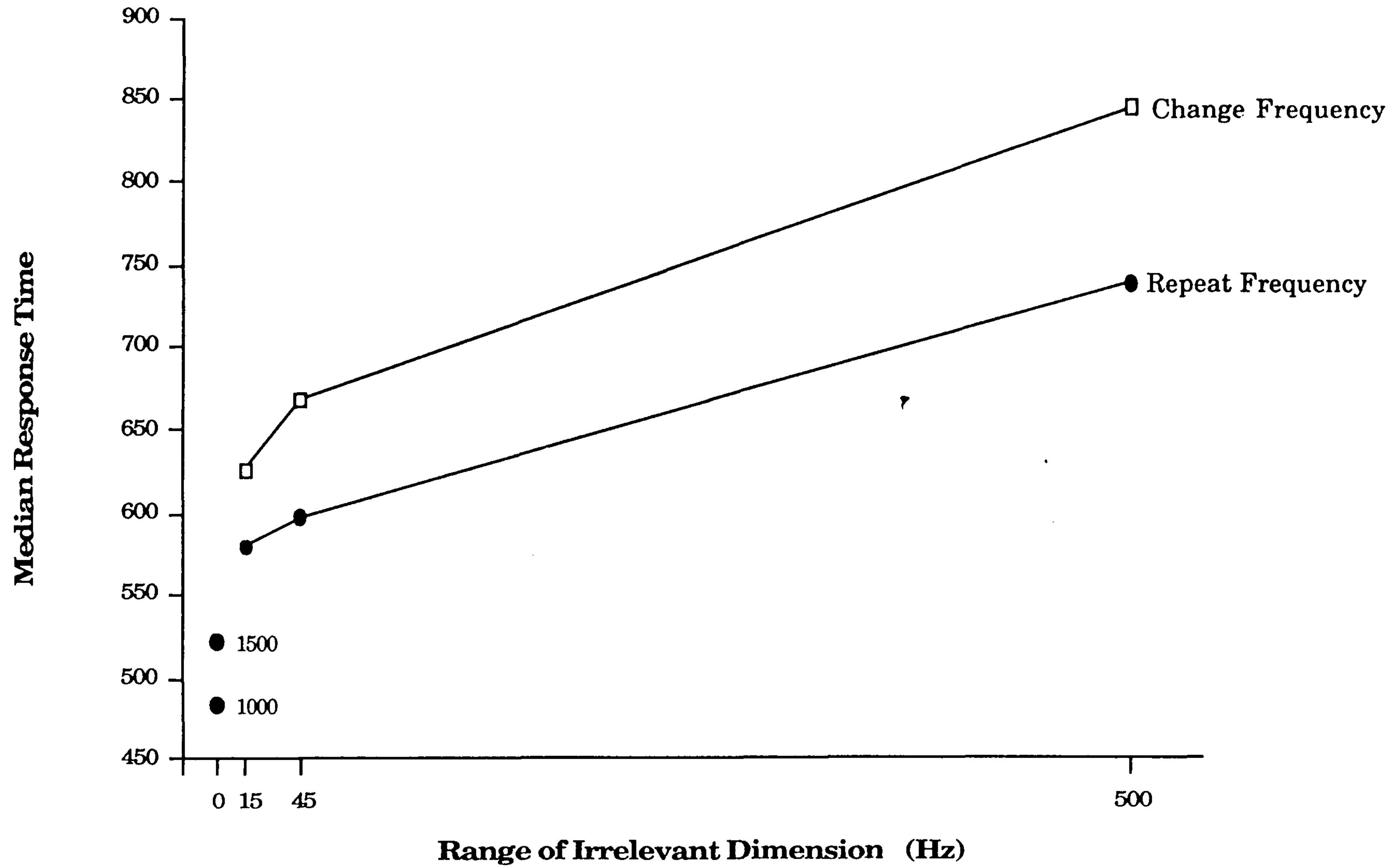

Response Time
2.4
2.2
2.0
1.8
1.6
1.4
0
1
2
3
4
5
6
7
8
9
10
Euclidean Distance between S N and S N - 1

Figure 6


Median Response Time
900
850
800
750
700
650
600
550
500
450
0 15 45
500
Range of Irrelevant Dimension (Hz)
Change Frequency
Repeat Frequency
1500
1000